

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2003, 00:24:04 / Search time 1393.27 Seconds  
(without alignments)  
9647.960 Million cell updates/sec

Title: US-09-763-902b-10

Perfect score: 830

Sequence: 1 tgaacttgcatcactg.....ataattaataacacatg 830

Scoring table: IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database:

EST:\*

1: em\_estba:\*  
2: em\_estba:\*  
3: em\_estba:\*  
4: em\_estba:\*  
5: em\_estba:\*  
6: em\_estba:\*  
7: em\_estba:\*  
8: em\_estba:\*  
9: em\_estba:\*  
10: em\_estba:\*  
11: em\_estba:\*  
12: em\_estba:\*  
13: em\_estba:\*  
14: em\_estba:\*  
15: em\_estba:\*  
16: em\_estba:\*  
17: em\_estba:\*  
18: em\_estba:\*  
19: em\_estba:\*  
20: em\_estba:\*  
21: em\_estba:\*  
22: em\_estba:\*  
23: em\_estba:\*  
24: em\_estba:\*  
25: em\_estba:\*  
26: em\_estba:\*  
27: em\_estba:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	766	92.3	982	9	AL551859
2	678	81.7	814	9	AL575327
3	645	77.7	685	14	BM975482
4	632.8	76.2	1110	13	BM543865
5	631.8	76.1	720	12	BF195575
6	629.4	75.8	631	10	AM274482

7	623	75.1	649	14	BM789812	BM789812 K-EST0069
8	613	73.9	644	14	BO581596	BO581596 110605.x
9	603	72.4	906	12	BG109830	BG109830 602280881
10	602	72.5	639	10	BE615929	BE615929 601279421
11	601.4	72.5	1125	14	BQ421087	BQ421087 AGENCOURT
12	592.4	71.4	627	10	BE615173	BE615173 601281104
13	591.8	71.3	882	12	BE794429	BE794429 601589445
14	590.6	71.2	606	10	BE643515	BE643515 h42c08.x
15	589.4	71.0	596	10	BE616032	BE616032 601279506
16	578.4	69.7	1002	13	BM470251	BM470251 AGENCOURT
17	560.4	67.5	576	12	BF062685	BF062685 7h63h09.x
18	554	66.7	577	14	BM723690	BM723690 UT-E-EJ0
19	553	66.6	553	9	A1739032	A1739032 w14e01.x
20	552	66.5	577	14	BO188558	BO188558 UT-E-EJ1
21	549.6	66.2	973	14	BO889949	BO889949 AGENCOURT
22	541.2	65.2	846	10	BE515286	BE515286 601236082
23	540.4	65.1	571	14	BM681875	BM681875 UT-E-EJ0
24	532.8	64.2	545	9	AA694201	AA694201 z142c08.s
25	526	63.4	527	9	A1967926	A1967926 w14a03.x
26	525	63.3	552	14	BM675619	BM675619 UT-E-EJ1
27	508.6	61.3	526	9	A1479307	A1479307 tm27a03.x
28	506.4	61.0	534	10	AM204159	AM204159 UT-H-B11
29	503.4	60.7	1076	13	BM552305	BM552305 AGENCOURT
30	498.8	60.1	942	12	BE790503	BE790503 601582508
31	496.2	59.8	727	13	B1913135	B1913135 603179795
32	495	59.6	511	10	AM205842	AM205842 UT-H-B11
33	483	58.2	514	14	BQ186422	BQ186422 UT-E-EJ1
34	481.8	58.0	868	14	BO937662	BO937662 AGENCOURT
35	477.8	57.6	512	12	BF683948	BF683948 na31d07
36	476.8	57.4	1033	12	BF683948	BF683948 602140366
37	474	57.1	493	14	BQ185616	BQ185616 UT-E-EJ1
38	473.4	57.0	1043	12	BG249456	BG249456 602361476
39	471.8	56.8	1057	12	BE733692	BE733692 601569289
40	471	56.7	471	14	BQ187897	BQ187897 UT-E-EJ1
41	467.8	56.4	786	13	B1560420	B1560420 603254527
42	467.4	56.3	939	11	AK003254	AK003254 Mus muscu
43	463.8	55.9	468	9	A1697839	A1697839 w617h05.x
44	462.4	55.7	478	9	A1807220	A1807220 w138c01.x
45	462.4	55.7	897	14	BO682672	BO682672 AGENCOURT

## ALIGNMENTS

RESULT 1  
LOCUS AL551859  
DEFINITION AL551859 LTI\_NFL006\_PL2 Homo sapiens CDNA clone CS0D1061YJ24 5  
ACCESSION AL551859  
VERSION AL551859.1 GI:128990210  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1. (bases 1 to 982)  
AUTHORS Li, W.B., Gruber, C., Jesse, J., and Polayes, D.  
TITLE Full-length CDNA libraries and normalization  
JOURNAL Unpublished (2001)  
COMMENT Contact: Genoscope  
Genoscope - Centre National de Sequencage  
BP 191 91006 EVRY cedex - France  
Email: sequef@genoscope.cns.fr, Web: www.genoscope.cns.fr.  
location/Qualifiers  
1. 982  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0D1061YJ24"  
/clone\_lib="LTI\_NFL006\_PL2"  
/issue\_type="placenta"  
/note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand CDNA  
was primed with a NotI-oligo(dT) primer. Five prime end

enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371  
 Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com"

BASE COUNT 211 a 270 c 263 g 232 t 6 others  
 ORIGIN

Query Match 92.3% Score 766; DB 9; Length 982;  
 Best Local Similarity 99.7% Pred. No. 9,2e-202;  
 Matches 766; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

49 CCCCTGTTCCCAAGGAGAGAAATACCTGTGAGCCCTCTTCATGAAACCA 108  
 215 CCCCTGTTCCCAAGGAGAGAAATACCTGTGAGCCCTCTTCATGAAACCA 274  
 109 GATGCAATCTGGATTTCAAGACCTATGTGATCAGGCGTGCAGAGCTGAGAGAT 168  
 275 GATGCAATCTGGATTTCAAGACCTATGTGATCAGGCGTGCAGAGCTGAGAGAT 334  
 169 TGTCATGTCTACTACACACCATGATTAAGCGGCGGCTTGTCTCCGCTGTACAT 228  
 335 TGTCATGTCTACTACACACCATGATTAAGCGGCGGCTTGTCTCCGCTGTACAT 394  
 229 GGGCAGACGACCCCTGTGTGTAATGCAATCTGTTTCAGACAAGAAATCTTTGATA 288  
 395 GGGCAGACGACCCCTGTGTGTAATGCAATCTGTTTCAGACAAGAAATCTTTGATA 454  
 289 GTTTTGAATGTGCTTCCAGCGAGTTCCTCAATACGCGTGAGTACCTGACGCTCT 348  
 455 GTTTTGAATGTGCTTCCAGCGAGTTCCTCAATACGCGTGAGTACCTGACGCTCT 514  
 349 TCATGTGAAGCCACACAGCAGACCGGCTCTGTGTGATCTGTGATCACTGAA 408  
 515 TCATGTGAAGCCACACAGCAGACCGGCTCTGTGTGATCTGTGATCACTGAA 574  
 409 GTTTGAGGGGAAACAACAGGGGCTTCAACAGAACTTCATCTTCAACGCCAGGCTTC 468  
 575 GTTTGAGGGGAAACAACAGGGGCTTCAACAGAACTTCATCTTCAACGCCAGGCTTC 634  
 469 ACCCAGACACAGTGTGAAAGATCCCAAGTCTGCTTCCAGAGACTGGGCGAG 528  
 635 ACCCAGACACAGTGTGAAAGATCCCAAGTCTGCTTCCAGAGACTGGGCGAG 694  
 529 CTAGTGGGGGTGGCAGAGGCTCTTTGCTTCATTCAGCCCTAGCTGTAGAGAAATGCA 588  
 695 CTAGTGGGGGTGGCAGAGGCTCTTTGCTTCATTCAGCCCTAGCTGTAGAGAAATGCA 754  
 589 AACCTGACCTCTCAAGATGTGAGGAACAAGTTCATTTCTGTGTTGGGAGACACTG 648  
 755 AACCTGACCTCTCAAGATGTGAGGAACAAGTTCATTTCTGTGTTGGGAGACACTG 814  
 649 CAGACTCAGCTGCCAGGTGAACTCTTTTGTGCTCAAGTGTAGAGAGCTCTT 708  
 815 CAGACTCAGCTGCCAGGTGAACTCTTTTGTGCTCAAGTGTAGAGAGCTCTT 874  
 709 CCTGAATATATATCTTTTGTGATAGTTCCTTTTCAAGTAGTAACTTTTATTTT 768  
 875 CCTGAATATATATCTTTTGTGATAGTTCCTTTTCAAGTAGTAACTTTTATTTT 934  
 769 CTACTTCCAGAGAGACTGTGATCTGTGAAATTTGACAAATAAT 816  
 935 CTACTTCCAGAGAGACTGTGATCTGTGAAATTTGACAAATAAT 982

RESULT 2 814 bp. mRNA linear EST 16-FEB-2001  
 AL575327 LOCUS AL575327 LTI.NFL006.PL2 Homo sapiens cDNA clone CS0D1061YJ24 3  
 DEFINITION prime, mRNA sequence.

ACCESSION AL575327  
 VERSION AL575327.1 GI:12936386  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo;  
 (bases 1 to 814).  
 AUTHOR Li, W.B., Gruber, C., Jesses, J., and Polayes, D.  
 TITLE Full-length cDNA libraries and normalization  
 JOURNAL Unpublished (2001)  
 COMMENT Contact: Genoscope  
 Genoscope - Centre National de Sequencage  
 BP 191 91006 Evry cedex - France  
 Email: segref@genoscope.cns.fr, Web : www.genoscope.cns.fr.

FEATURES  
 source  
 1. 814  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CS0D1061YJ24"  
 /clone\_1b="LTI.NFL006.PL2"  
 /tissue\_type="Placenta"  
 /note="Vector: pCMVSPORT 6; Site 1: NotI; 1st strand cDNA was primed with a NotI-oligo(dT) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies, a division of Invitrogen 9800 Medical Center Drive, Rockville, Maryland 20850, USA Fax : (1) 301 610 8371  
 Email : fliang@lifetech.com URL :  
 http://fulllength.invitrogen.com"

BASE COUNT 190 a 219 c 215 g 174 t 16 others  
 ORIGIN

Query Match 81.7% Score 678; DB 9; Length 814;  
 Best Local Similarity 96.3% Pred. No. 2.4e-177;  
 Matches 697; Conservative 15; Mismatches 10; Indels 2; Gaps 2;

50 CCCTGTTCCCAAGGAGAGAAATACCTGTGAGCCCTCTTCATGAAACCA 109  
 723 CCCTGTTCCCAAGGAGAGAAATACCTGTGAGCCCTCTTCATGAAACCA 664  
 110 ATGCAATCTGTGATTTCAAGACCTATGTGATCAGGCGTGCAGAGCTGTGAGAGATT 169  
 663 ATGCAATCTGTGATTTCAAGACCTATGTGATCAGGCGTGCAGAGCTGTGAGAGATT 604  
 170 GTCATGTCTACTACACACCATGATTAAGCGGCGGCTTGTGCTCCGCTGTACATG 229  
 603 GTCATGTCTACTACACACCATGATTAAGCGGCGGCTTGTGCTCCGCTGTACATG 545  
 230 GGCACAGCACCCTGTGTGTAATGCAATCTGTTTCAAGACAAGATCCTTGAAGT 289  
 544 GGCACAGCACCCTGTGTGTAATGCAATCTGTTTCAAGACAAGATCCTTGAAGT 485  
 290 TTTTGAATGTGCTTCCAGCGAGTTCCTTCAAGTACGCTGTGAGAGCTGCACTGTT 349  
 484 TTTTGAATGTGCTTCCAGCGAGTTCCTTCAAGTACGCTGTGAGAGCTGCACTGTT 425  
 350 CATATGAAGCCACACAGCGAGCAGCG -TCCTGTTGTCAATCTGTGATCAGTAA 408  
 424 CATATGAAGCCACACAGCGAGCAGCG -TCCTGTTGTCAATCTGTGATCAGTAA 365  
 409 GTTTGAGGGGAAACAACAGGGGCTTCAACAGAACTTCATCTGACGCCAGGCGCTC 468  
 364 GTTTGAGGGGAAACAACAGGGGCTTCAACAGAACTTCATCTGACGCCAGGCGCTC 305  
 469 ACCCAGACACAGTGTGAGAGATGCAAGTACTGCTTCCGCTTCCAGAGACTGGGCGAG 528  
 304 ACCCAGACACAGTGTGAGAGATGCAAGTACTGCTTCCGCTTCCAGAGACTGGGCGAG 245  
 529 CTAGTGGGGGTGGCAGAGGCTCTTTGCTTCATTCAGCCCTAGCTGTAGAGAAATGCA 588



Db 244 CTAGTGGGGGAGAGAGAGTCTTGGCTGATTCAGCCCTAGCTCTGTAGAAACCA 185  
 Oy 589 AACCTGCAGCTCAGAGATGTAGAGAACAAAGTTCATTTCTGTGTGGAGAGACTG 648  
 Db 184 AACCTGCAGCTCAGAGATGTAGAGAACAAAGTTCATTTCTGTGTGGAGAGACTG 125  
 Oy 649 CAGAGCTCAGCTCAGAGAGTGTAGAGTTCATTTCTGTGTGGAGAGACTG 708  
 Db 124 CAGAGCTCAGCTCAGAGAGTGTAGAGTTCATTTCTGTGTGGAGAGACTG 65  
 Oy 709 CCTGAATATATATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 768  
 Db 64 CCTGAATATATATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 5  
 Oy 769 CTAC 772  
 Db 4 CGAC 1  
 RESULT 3  
 BM975482/C  
 LOCUS BM975482 685 bp mRNA linear EST 21-MAR-2002  
 DEFINITION UI-CF-EN1-acv-k-12-0-UI.s1 UI-CF-EN1 Homo sapiens cDNA clone  
 BM975482  
 UI-CF-EN1-acv-k-12-0-UI 3', mRNA sequence.  
 ACCESSION BM975482.1 GI:19593073  
 VERSION EST.  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 685)  
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.  
 TITLE Normalization and subtraction: two approaches to facilitate gene  
 discovery  
 JOURNAL Genome Res. 6 (9), 791-806 (1996)  
 MEDLINE 9704447  
 COMMENT Contact: McCray, PB  
 McCray Lab  
 University of Iowa  
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA  
 Tel: 319 356 4866  
 Fax: 319 356 7171  
 Email: paul-mccray@iowa.edu  
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa  
 cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa  
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa  
 Clone Distribution: Researchers may obtain clones from Research  
 Genetics (www.resgen.com).  
 The following repetitive elements were found in this cDNA  
 sequence: 1-27, >AT-rich#low-complexity (matched complement)  
 Seq primer: M13 FORWARD  
 POLY-A=yes:  
 FEATURES  
 source  
 1. 685  
 Location/Qualifiers  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="UI-CF-EN1-acv-k-12-0-UI"  
 /clone\_lib="UI-CF-EN1"  
 /tissue\_type="Primary Lung Cystic Fibrosis Epithelial  
 Cells"  
 /dev\_stage="Adult"  
 /lab\_host="DH10B (Life Technologies) (T1 phage resistant)"  
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a  
 modified polylinker; Site 1: EcoR I; Site 2: Not I;  
 UI-CF-EN1 is a normalized cDNA library containing the  
 following tissue(s): Primary Lung Cystic Fibrosis  
 Epithelial Cells. The library was constructed according to  
 Bonaldo, Lennon and Soares, Genome Research, 6:791-806,  
 1996. First strand cDNA synthesis was primed with an  
 oligo-dT primer containing a Not I site. Double stranded  
 cDNA was ligated to an EcoR I adaptor, digested with Not

I, and cloned directionally into pT73-Pac vector. The  
 oligonucleotide used to prime the synthesis of  
 first-strand cDNA contains a library tag sequence that is  
 located between the Not I site and the (AT)18 tail. The  
 sequence tag for this library is CTGCTCAGGT.  
 TAG\_LIB-UI-CF-EN1  
 TAG\_TISSUE-Human Lung Epithelial Cell Lines untreated LPS  
 6hr to LPS 24h  
 TAG\_SEQ-CTGCTCAGGT  
 BASE COUNT 184 a 164 c 163 g 174 t  
 ORIGIN  
 Query Match 77.7%; Score 645; DB 14; Length 685;  
 Best Local Similarity 99.7%; Pred. No. 3.4e-168;  
 Matches 667; Conservative 0; Mismatches 0; Indels 2; Gaps 2;  
 Oy 104 CCAGAGATGCGATCTGTGGATTTCAGAGCTATGTGATCGAGGCTCGAGAGCTGAG 163  
 Db 685 CCAGAGATGCGATCTGTGGATTTCAGAGCTATGTGATCGAGGCTCGAGAGCTGAG 626  
 Oy 164 GAGTTGTCAATGTCTACTACACCAACATGATAGCGGGCGGCTTGTGTCGCCGCTG 223  
 Db 625 GAGTTGTCAATGTCTACTACACCAACATGATAGCGGGCGGCTTGTGTCGCCGCTG 566  
 Oy 224 TACATGGGACAGCCACCGCTGTGGAATGCAATGCTTTCAGAGAACAAATCCTTG 283  
 Db 565 TACATGGGACAGCCACCGCTGTGGAATGCAATGCTTTCAGAGAACAAATCCTTG 506  
 Oy 284 AGTGAGTTTTCGAATGTGCTTCAGAGGATTCGAATGAGGCTGTAGACTGCCAG 343  
 Db 505 AGTGAGTTTTCGAATGTGCTTCAGAGGATTCGAATGAGGCTGTAGACTGCCAG 447  
 Oy 344 CCTGTCATGATGAAGCCACACCAAGCCAGACCGGCTCTTGTGTGTCATGTGGATCA 403  
 Db 446 CCTGTCATGATGAAGCCACACCAAGCCAGACCGGCTCTTGTGTGTCATGTGGATCA 387  
 Oy 404 GTGAGTTGTAGGGAGAACAAAGGGAGCTTCAACGAAGTCACTGACCGCCAG 463  
 Db 386 GTGAGTTGTAGGGAGAACAAAGGGAGCTTCAACGAAGTCACTGACCGCCAG 327  
 Oy 464 GCCTACCCAGACACACAGTGTGAAGATGCAAGTCACTGCTCCGCTCCAGAGACTG 523  
 Db 326 GCCTACCCAGACACACAGTGTGAAGATGCAAGTCACTGCTCCGCTCCAGAGACTG 267  
 Oy 524 GCCAGCTAGTGGGGGTGGCAGAGTCTCTTCTTCATTCACGCTAGCTGTAGAGAA 583  
 Db 266 GCCAGCTAGTGGGGGTGGCAGAGTCTCTTCTTCATTCACGCTAGCTGTAGAGAA 207  
 Oy 584 ATGCAAACTGACCTCAAGATGTGAGGAGACAAAGTTCATTTCTGTGGGGAGA 643  
 Db 206 ATGCAAACTGACCTCAAGATGTGAGGAGACAAAGTTCATTTCTGTGGGGAGA 147  
 Oy 644 CACTCAGAGCTCAGAGTGTGCGAGGTGAACTCTTTTGTGTCTCAAGTCTAGAGCT 703  
 Db 146 CACTCAGAGCTCAGAGTGTGCGAGGTGAACTCTTTTGTGTCTCAAGTCTAGAGCT 87  
 Oy 704 CCTTCTCGAATATATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 762  
 Db 86 CCTTCTCGAATATATCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 27  
 Oy 763 ATTTCCTCA 771  
 Db 26 ATTTCCTCA 18  
 RESULT 4  
 BM543865 1110 bp mRNA linear EST 20-FEB-2002  
 LOCUS BM543865  
 DEFINITION AGENCOU007\_6492315 NIH\_MGC\_125 Homo sapiens cDNA clone IMAGE:5589207  
 5', mRNA sequence.  
 ACCESSION BM543865  
 VERSION BM543865.1 GI:18774629  
 KEYWORDS EST.

SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 1110)  
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/  
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)  
COMMENT Unpublished (1999)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: Invitrogen  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LINL at:  
http://image.lnl.gov  
Plate: LHAM12361. row: 1 column: 16  
High quality sequence stop: 563.  
Location/Qualifiers  
1. 1110  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="5589207"  
/clone\_lib="NIH-MGC\_125"  
/lab\_host="DH10B"  
/note="Organ: ovary (pool of 3); Vector: pCMV-Sport6; Site\_1: EcoRV (destroyed); Site\_2: NotI; RNA source: pool of three ovaries, from females ranging in age from 38 to 49 yo. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 2.1 kb, insert size range 1.3-5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 036."

BASE COUNT 249 a 310 c 282 g 259 t 10 others  
ORIGIN

Query Match 76.28; Score 632.8; DB 13; Length 1110;  
Best Local Similarity 90.98; Pred. No. 9,7e-165;  
Matches 719; Conservative 0; Mismatches 63; Indels 9; Gaps 4;

41 CCCTGACGCCCCCTGCTCCCAAGGAGAGAAATACCTGTGAGACCTCTCCATA 100  
117 CCGCGACGCCCCCTGCTCCCAAGGAGAGAAATACCTGTGAGACCTCTCCATA 176  
101 GAACAGAGATGAGATCTGTGATTTCAAGACCTATGTGATCAGGCTGACAGAGTGT 160  
177 GAACAGAGATGAGATCTGTGATTTCAAGACCTATGTGATCAGGCTGACAGAGTGT 236  
161 GAGGAGTTTCTCATGCTCTACACACCATGATAGGGGGGGGCTTGTCTCCCGC 220  
237 GAGGAGTTTCTCATGCTCTACACACCATGATAGGGGGGGGCTTGTCTCCCGC 296  
221 CTGTACATGGGACAGGACCCCTGCTGTGATGAGGACCTCTGTTCAGAGACAAGATCC 280  
297 CTGTACATGGGACAGGACCCCTGCTGTGATGAGGACCTCTGTTCAGAGACAAGATCC 356  
281 TTGAGTGTGAGTTTGTGAATGTGCTTCAGGAGAGTTCAGGAGTGTAGACTGC 340  
357 TTGAGTGTGAGTTTGTGAATGTGCTTCAGGAGAGTTCAGGAGTGTAGACTGC 416  
341 CAGGCTGTATATGAGGACCAACAGGAGCTTCAACAGAGACTTCATCTGACCGCC 460  
417 CAGGCTGTATATGAGGACCAACAGGAGCTTCAACAGAGACTTCATCTGACCGCC 476  
401 TCAGTGAAGTTTGAAGGACCAACAGGAGCTTCAACAGAGACTTCATCTGACCGCC 460  
477 TCAGTGAAGTTTGAAGGACCAACAGGAGCTTCAACAGAGACTTCATCTGACCGCC 536  
461 CAGGCTGTATATGAGGACCAACAGGAGCTTCAACAGAGACTTCATCTGACCGCC 520  
537 CAGGCTGTATATGAGGACCAACAGGAGCTTCAACAGAGACTTCATCTGACCGCC 596

521 TGGCCAGCTAGTGGGGGTGGCAGAGGTCCTTGTCTATTCAGCCCTAGCTGTAGA 580  
597 TGGCCAGCTAGTGGGGGTGGCAGAGGTCCTTGTCTATTCAGCCCTAGCTGTAGA 654  
581 GAATGCAAAACCTGCACTCTCAAGATGTGAGAGACAGCAAGTTCATCTGTGTGGCG 640  
655 GAATGCAAAACCTGCACTCTCAAGATGTGAGAGACAGCAAGTTCATCTGTGTGGCG 714  
641 AGCACTGCAAGATCCACATGTGCGGAGGTGAAGTCTTTTGTGCTCAAGTTCAAGA 700  
715 GAACACTGCAAGATCCACATGTGCGGAGGTGAAGTCTTTTGTGCTCAAGTTCAAGA 774  
701 GTCCCTTCTGCAATATATCTTCTTGTGCTCAAGTTCAGAGTTCAGAACTT 758  
775 GTCCCTTCTGCAATATATCTTCTTGTGCTCAAGTTCAGAGTTCAGAACTT 834  
759 TT---CTATTTTCTACTTGGCCAGTACAGACTC--TGATTCGGAATTCAGCAATA 813  
835 TTTCATATTTTTCACACTGGCCAGTAAATAATCCTGAGATCTGGAATAATCTGACAAAT 894  
814 ATTTAATATA 824  
895 AATTTAATATA 905

RESULT 5  
BF195575/c 720 bp mRNA linear EST 03-NOV-2000  
LOCUS 7n85b11.x1 NCI-CGAP\_Ov18 Homo sapiens CDNA clone IMAGE:3571173.3  
DEFINITION similar to TR:Q9UKK6 Q9UKK6 NTF2-RELATED EXPORT PROTEIN NXY1.1;  
mRNA sequence.  
BF195575  
BF195575.1 GI:11082605  
EST.  
KEYWORDS  
VERSION  
ACCESSION  
ORIGIN  
SOURCE  
ORGANISM human.  
Homo sapiens  
Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 720)  
NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
Tumor Gene Index  
Unpublished (1997)  
Contact: Robert Strausberg, Ph.D.  
Email: cgaabs-remail.nih.gov  
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D. CDNA Library Preparation: M. Bento Soares, Ph.D. CDNA Library Arrayed by: Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by: Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LINL, send email to:  
info@image.lnl.gov  
High quality sequence stop: 466.  
Location/Qualifiers  
1. 720  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_image="3571173"  
/clone\_lib="NCI CGAP Ov18"  
/tissue\_type="fibrocyte"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: ovary; Vector: pTV73D-Pac (Pharmacia) with a modified polylinker; Site\_1: NotI; Site\_2: EcoRI; 1st strand cDNA was primed with a NotI - 'oligo(dT) primer [5' TGTTACCAATCTGAGAGGAGGCGGCGGCGGCAATTTTTTTTTTTT 3']; double-stranded cDNA was ligated to EcoRI adaptors (Pharmacia), digested with NotI and cloned into the NotI and EcoRI sites of the modified pTV73 vector. Library went through one round of normalization, and was constructed by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 205 a 170 c 169 g 175 t 1 others



Db 31 TTTCAAGTAGTAACCTTTCTATTTCCT 1

RESULT 7  
LOCUS BM789812 649 bp mRNA linear EST 05-MAR-2002  
DEFINITION K-EST0069490 S22SNU16 Homo sapiens cDNA clone S22SNU16-1-G02 5'

ACCESSION BM789812  
VERSION BM789812.1 GI:19138044  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 649)  
Kim, N.S., Hahn, Y., Oh, J.H., Lee, J.Y., Ahn, H.Y., Chu, M.Y., Kim, M.R.,  
Oh, K.J., Cheong, J.E., Sohn, H.Y., Kim, J.M., Park, H.S., Kim, S. and:  
Kim, Y.S.  
21C Frontier Korean EST Project 2001  
Unpublished (2002)  
Contact: Kim YS  
Genome Research Center  
Korea Research Institute of Bioscience & Biotechnology  
52 Eoeun-dong Yuseong-gu, Daejeon 305-333, South Korea  
Tel: +82-42-860-4470  
Fax: +82-42-860-4409  
Email: Yongsung@mail.kribb.re.kr  
Plate: 1 row: 6 column: 02  
High quality sequence stop: 649.  
Location/Qualifiers  
1. 649  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="S22SNU16-1-G02"  
/clone\_1lb="S22SNU16"  
/sex="F"  
/tissue\_type="Ascites"  
/cell\_type="Lymphoblast-like"  
/cell\_line="SNU-16"  
/lab\_host="DH10B"  
/note="Organ: Stomach; Vector: pTT3-Pac; Site: 1; EcoRI;  
Site\_2: NotI; The S22SNU16 library was contributed by the  
Soares laboratory and it was constructed as described by  
Bonald, M.F., Lennon, G. and Soares, M.B. (1996), Genome  
Research 6(9): 791-806. RNA was prepared from harvested  
cells of SNU-16 culture. SNU-16 cell was obtained from  
Korean Cell Line Bank (KCLB). SNU-16 was established from  
ascitic fluids of Korean patients by Park J.G. et al.  
(1990), Cancer Res 50: 2773-2780."

BASE COUNT 156 a 171 c 172 g 150 t

ORIGIN

Query Match 75.1%; Score 623; DB 14; Length 649;  
Best Local Similarity 100.0%; Pred. No. 4.4e-162;  
Matches 623; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

49 CCCCCTGTTCCCAAGGAGGAGGAATACCTGTGAGGACCTCTCCATAGAACCGA 108  
DB 27 CCCCCTGTTCCCAAGGAGGAGGAATACCTGTGAGGACCTCTCCATAGAACCGA 86  
109 GATGCACTGTGATTTCAAGACCTATGTGATCAGGCTGACAGAGCTGTGAGAGTT 168  
DB 87 GATGCACTGTGATTTCAAGACCTATGTGATCAGGCTGACAGAGCTGTGAGAGTT 146  
169 TGTCAATGTCTACTACACACCATGATTAAGGGGGGGTGTGCTCCCGCTTACAT 228  
DB 147 TGTCAATGTCTACTACACACCATGATTAAGGGGGGGTGTGCTCCCGCTTACAT 206  
229 GGGCAGGACACCGGTCTGGAATGCAATCTTTTCAGACACAAGATCCTTGAGTGA 288  
DB 207 GGGCAGGACACCGGTCTGGAATGCAATCTTTTCAGACACAAGATCCTTGAGTGA 266

QY 289 GTTTTGAATGTTCCCTCCAGCAGAGTTCGAATACAGCTGTAGACTGCCAGCTGT 348  
DB 267 GTTTTGAATGTTCCCTCCAGCAGAGTTCGAATACAGCTGTAGACTGCCAGCTGT 326  
QY 349 TCATGATGAGCCACACCAAGCCAGACGAGCTGTGTTGATCTGTGAGTGA 408  
DB 327 TCATGATGAGCCACACCAAGCCAGACGAGCTGTGTTGATCTGTGAGTGA 386  
QY 409 GTTTGAGGGGACAAACACAGGAGCTTCAACAGAACTTCATCCAGCCAGGCTC 468  
DB 387 GTTTGAGGGGACAAACACAGGAGCTTCAACAGAACTTCATCCAGCCAGGCTC 446  
QY 469 ACCCAGACACACAGTGTGAGATCCAGAGTGAAGTCTGCTCCAGTCCAGAGTGGCCAG 528  
DB 447 ACCCAGACACACAGTGTGAGATCCAGAGTGAAGTCTGCTCCAGTCCAGAGTGGCCAG 506  
QY 529 CTAGTGGGGGTGACAGAGTCTCTTCTTCATTCACAGCCCTAGCTGTGAGAAATGCA 588  
DB 507 CTAGTGGGGGTGACAGAGTCTCTTCTTCATTCACAGCCCTAGCTGTGAGAAATGCA 566  
QY 589 AACCTGACTCTCAAGATGTGAGAGACACAAAGTCAATTTCTGTTGGGGAGACACTG 648  
DB 567 AACCTGACTCTCAAGATGTGAGAGACACAAAGTCAATTTCTGTTGGGGAGACACTG 626  
QY 649 CAGACTCCACTGTGCGGAGTTG 671  
DB 627 CAGACTCCACTGTGCGGAGTTG 649

RESULT 8  
LOCUS B0581596/ 644 bp mRNA linear EST 20-JUN-2002  
DEFINITION 110905.x1 Human insulinoma Homo sapiens cDNA clone IMAGE:6029745  
3' similar to IR:Q9UKK6 Q9UKK6 NIF2-RELATED EXPORT PROTEIN N1X1.  
mRNA sequence.  
B0581596  
B0581596.1 GI:21494492  
EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 644)  
Lemstra, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,  
Lemstra, D., Scaer, M., Brestelli, J., Gradwohl, G., Clifton, S.,  
Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blisstein, A.,  
Schmitt, A., Theising, B., Rittler, E., Ronko, T., Bennett, J., Gardenas,  
M., Gibbons, M., McCann, R., Cole, R., Tsagarelis, R., Williams, T.,  
Jackson, Y. and Bowers, Y.  
Endocrine Pancreas Consortium  
Unpublished (2000)  
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Harvard University, Howard Hughes Medical Institute  
Dept. of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@hms.harvard.edu  
Library was constructed by Dr. J. Ferrer in vivo mass-excised to  
pBluescript SK- by Dr. H. Inoue DNA sequencing by Washington  
University Genome Sequencing Center for information on obtaining a  
Seq primer: -400P from Gibco  
High quality sequence stop: 488.  
Location/Qualifiers  
1. 644  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone\_1lb="Human Insulinoma"  
/tissue\_type="Insulinoma"

TITLE  
JOURNAL  
COMMENT

/lab.host="DH10B (phage-resistant)"  
 /note="Organ: pancreas; Vector: pBluescript SK-; Site: 1;  
 XhoI; Site: 2; EcoRI; Constructed with lambda ZapII system  
 (Stratagene) by Dr. J. Ferrier, in vivo mass-excised to  
 pBluescript SK- by Dr. H. Inoue following the Washington  
 University protocol  
 (http://genome.wustl.edu/est/lambda\_protocol.shtml).  
 Please contact Hiroshi Inoue, MD/PhD for further  
 information on this library (Metabolism Division, Permutt  
 Laboratory, Washington University School of Medicine, Box  
 8127, 660 S Euclid Ave, St. Louis, MO 63110). Note: this  
 is a Washington University Pancreas EST project library."

BASE COUNT 179 a 142 c 149 g 174 t  
 ORIGIN

Query Match 73.9%; Score 613; DB 14; Length 644;  
 Best Local Similarity 100.0%; Pred. No. 2.6e-159;  
 Matches 613; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

218 GCGCTGATACATGAGGACACAGCCCTGCTGCTGATGCAATGCTTTCCAGACAGAA 277  
 644 GCGCTGATACATGAGGACACAGCCCTGCTGCTGATGCAATGCTTTCCAGACAGAA 585  
 278 TCCCTGAGTGAAGTTTGAATGTTGCTTCCAGAGCTTCCAAATCAGCTGTAGAC 337  
 584 TCCCTGAGTGAAGTTTGAATGTTGCTTCCAGAGCTTCCAAATCAGCTGTAGAC 525  
 338 TCCAGAGCTGTTATGATGATGACCAACACAGCCAGACAGCTGCTGTGTATCTGT 397  
 524 TCCAGAGCTGTTATGATGATGACCAACACAGCCAGACAGCTGCTGTGTATCTGT 465  
 398 GGATCAGTAAATTTGAGGGGGAACAACAAGGGGATTTACCAAGACTTCACTCCGACC 457  
 464 GGATCAGTAAATTTGAGGGGGAACAACAAGGGGATTTACCAAGACTTCACTCCGACC 405  
 458 GCGCAGGCTCAGCCAGCAACAGATGTTGGAAGATGCAAGTACTGCTCCGCTCCAG 517  
 404 GCGCAGGCTCAGCCAGCAACAGATGTTGGAAGATGCAAGTACTGCTCCGCTCCAG 345  
 518 GACTGGGCGAGTGAAGTGGGGGTGGGAGAGTCTCTTTGCTTCAATGAGCCCTGCTGT 577  
 344 GACTGGGCGAGTGAAGTGGGGGTGGGAGAGTCTCTTTGCTTCAATGAGCCCTGCTGT 285  
 578 AGAAGAAATGCAAACTGCACTCTCAAGATGTGAGGAACAAGTTCATTTCTGTGTG 637  
 284 AGAAGAAATGCAAACTGCACTCTCAAGATGTGAGGAACAAGTTCATTTCTGTGTG 225  
 638 CGAGAGACTGAGCACTGCACTGTCGAGGTTGAATCTTTTGTGTGCTCAAGTTCTTA 697  
 224 CGAGAGACTGAGCACTGCACTGTCGAGGTTGAATCTTTTGTGTGCTCAAGTTCTTA 165  
 698 GAGTCCCTTCTCTGATATATATCTGTTGTCATAGTTTCTTTCAAGTAGTAAGT 757  
 164 GAGTCCCTTCTCTGATATATATCTGTTGTCATAGTTTCTTTCAAGTAGTAAGT 105  
 758 TTTCTATTTTCTACTTGGCCAGTACAGACTCTGATTTGGAATTTCTGCAATAATTT 817  
 104 TTTCTATTTTCTACTTGGCCAGTACAGACTCTGATTTGGAATTTCTGCAATAATTT 45  
 818 AATAATACACATG 830  
 44 AATAATACACATG 32

RESULT 9  
 BG109830 906 bp mRNA linear EST 30-JAN-2001  
 LOCUS 602280989F1 NIH\_MGC\_86 Homo sapiens cDNA clone IMAGE:4368311 5'  
 DEFINITION mRNA sequence.  
 ACCESSION BG109830  
 VERSION BG109830.1 GI:12603336  
 KEYWORDS EST.  
 SOURCE human.

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 REFERENCE 1 (bases 1 to 906)  
 AUTHORS NIH-MGC http://mhc.nci.nih.gov/  
 TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cga@bbs-remail.nih.gov  
 Tissue Procurement: ATCC  
 Tissue Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
 http://image.llnl.gov  
 Plate: LLNL0022 row: b column: 24  
 High quality sequence stop: 705.  
 Location/Qualifiers

FEATURES  
 source 1..906  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:4368311"  
 /clone\_lib="NIH\_MGC\_86"  
 /tissue\_type="osteosarcoma, cell line"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: bone; Vector: pCMV-Sport6; Site: 1; NotI;  
 Site: 2; SalI; Cloned unidirectionally; oligo-dT primed.  
 Average insert size 1.53 kb. Library enriched for  
 full-length clones and constructed by Life Technologies.  
 Note: this is a NIH-MGC library"

BASE COUNT 197 a 252 c 248 g 208 t 1 others  
 ORIGIN

Query Match 73.4%; Score 609; DB 12; Length 906;  
 Best Local Similarity 93.8%; Pred. No. 3.7e-158;  
 Matches 678; Conservative 0; Mismatches 40; Indels 5; Gaps 4;

41 CCCTGAGCCCTGCTTCCCAAGGACAGAAATACCTGGGAGCCCTCTCCATA 100  
 118 CCGCGAGCCCTGCTTCCCAAGGACAGAAATACCTGGGAGCCCTCTCCATA 177  
 101 GAACAGAGATGATCTGTGATTTCAAGACCTATGTGATGATGAGCCCTCAGAGCTGT 160  
 178 GAACAGAGATGATCTGTGATTTCAAGACCTATGTGATGATGAGCCCTCAGAGCTGT 237  
 161 GAGAGATTTGTCAATGCTCTACACACACCATGATTAAGCGCGCTTGTGTCTGCCGC 220  
 238 GAGAGATTTGTCAATGCTCTACACACACCATGATTAAGCGCGCTTGTGTCTGCCGC 297  
 221 CTGTACATGGGACAGCCCTGCTGTGATGATGATGATGATGATGATGATGATGATG 280  
 298 CTGTACATGGGACAGCCCTGCTGTGATGATGATGATGATGATGATGATGATGATG 357  
 281 TTGAGTGAAGTTTGAATGTTGCTTCCAGGAGTTTCAAAATCAGCGTGAAGATGC 340  
 358 TTGAGTGAAGTTTGAATGTTGCTTCCAGGAGTTTCAAAATCAGCGTGAAGATGC 417  
 418 CAGCCTGTTATGATGATGAAGCAACACCAAGCCAGCGGCTTGTGTCTGTGTGA 477  
 401 TCACTGAAGTTTGAAGGGAACAACAAGGAGCTTCAACAGAACTTCACTGACCGCC 460  
 478 TCACTGAAGTTTGAAGGGAACAACAAGGAGCTTCAACAGAACTTCACTGACCGCC 537  
 461 CAGCCTGTTATGATGATGAAGCAACACCAAGCCAGCGGCTTGTGTCTGTGTGA 400  
 418 CAGCCTGTTATGATGATGAAGCAACACCAAGCCAGCGGCTTGTGTCTGTGTGA 477  
 401 TCACTGAAGTTTGAAGGGAACAACAAGGAGCTTCAACAGAACTTCACTGACCGCC 460  
 478 TCACTGAAGTTTGAAGGGAACAACAAGGAGCTTCAACAGAACTTCACTGACCGCC 537  
 521 TGGGCGAGCTAGTGGGGGTGGCAGAGGCTTC-TTGGCTTCACTCAGCCCTAGCTGTAG 579  
 598 TGGGCGAGCTAGTGGGGGTGGCAGAGGCTTC-TTGGCTTCACTCAGCCCTAGCTGTAG 655

DB 580 AGAATGCAAACTGACCTCTCAAGATGTGAGAA-CACAAGTTCATTTCTGTGTC 638  
DB 656 AGAAAGCAAACTGACCTCTCAAGATGTGAGAA-CACAAGTTCATTTCTGTGTC 715  
DB 639 GGAGACACTGACG-AGTCACTGTGCGGAGGTGAGACTCTTTTGTGCTCAAGTTCA 697  
DB 716 GGAGACACTGACGAACTCCACTGCGCGGAGGTGAGACTCTTTTGTGCTCAAGTTCA 775  
DB 698 GGAGCTCTCTCTCTGTAATATATATCTGTTGTCATAGTTCTTTCAAGTAGTAAGT 757  
DB 776 GGAGCTCTCTCTCTGTAATATATATCTGTTGTCAGACAAGTTCCCTTCCAGTAACTTCT 835  
DB 758 TTT 760  
DB 836 TTT 838

RESULT 10  
LOCUS BE615929 639 bp mRNA linear EST 24-AUG-2000  
DEFINITION 601279421F1 NIH\_MGC\_39 Homo sapiens cDNA clone IMAGE:3611596 5',  
mRNA sequence.  
ACCESSION BE615929  
VERSION BE615929.1 GI:9897615  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 639)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: ATCC  
CDNA Library Preparation: Ling Hong/Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM269 row: 1 column: 05  
High quality sequence stop: 639.  
Location/Qualifiers  
1. 639  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:3611596"  
/clone\_lib="NIH\_MGC\_39"  
/tissue\_type="adenocarcinoma"  
/lab\_host="DH10B (phage-resistant)"  
/note="Organ: pancreas; Vector: pOTB7; Site:1: XhoI;  
Site:2: EcoRI; cDNA made by oligo-dT priming.  
Directionally cloned into EcoRI/XhoI sites using the  
following 5' adaptor: GGCACGAG(G). Library constructed  
by Ling Hong in the laboratory of Gerald M. Rubin  
(University of California, Berkeley) using ZAP-cDNA  
synthesis kit (Stratagene) and Superscript II RT (Life  
Technologies)."

BASE COUNT 148 a 158 c 179 t  
ORIGIN

Query Match 72.5%; Score 602; DB 10; Length 639;  
Best Local Similarity 99.2%; Pred. No. 3e-156;  
Matches 605; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

DB 140 GATCAGGCTCTGACAGCTGCTGAGAGTTTGTCAATGTCTACTACACCATGATAG 199  
1 GATCAGGCTCTGACAGCTGCTGAGAGTTTGTCAATGTCTACTACACCATGATAG 60  
DB 200 CGGCGGCTTTCTGTCTGCTGCTGATGAGGACACGACCCCTGTCTGATGCAAT 259

DB 61 CGGCGGCTTTCTGTCTGCTGCTGCTGATGAGGACACGACCCCTGTGATGCAAT 120  
DB 260 GCTGTTTCAGACAAATCTCTGAGTGAATTTTGAATGTGCTTCACAGAGTTG 319  
DB 121 GCTGTTTCAGACAAATCTCTGAGTGAATTTTGAATGTGCTTCACAGAGTTG 180  
DB 320 CAATAGCGGTGTAGACTGCCAGCTGTCATGATGAAACCAACCAACCAACGAC 379  
DB 181 CAATAGCGGTGTAGACTGCCAGCTGTCATGATGAAACCAACCAACCAACGAC 240  
DB 380 GTCTCTGTTGTCTCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 439  
DB 241 GTCTCTGTTGTCTCTGTGATGATGATGATGATGATGATGATGATGATGATGAT 300  
DB 440 CAGACTCATCTGACCGCGCCAGGCTCACCAGCAACCACTGTGGAAGATGCAAGT 439  
DB 301 CAGACTCATCTGACCGCGCCAGGCTCACCAGCAACCACTGTGGAAGATGCAAGT 360  
DB 500 GACTGCTTCGCTTCAGAGACTGGCCAGCTAGTGGGGGTGCGAGAGTCTTTGCTTC 559  
DB 361 GACTGCTTCGCTTCAGAGACTGGCCAGCTAGTGGGGGTGCGAGAGTCTTTGCTTC 420  
DB 560 ATTACAGCCCTAGCTCTGTAGAGAAATGCAAACTGCTCAAGATGTGAGAAACA 619  
DB 421 ATTACAGCCCTAGCTCTGTAGAGAAATGCAAACTGCTCAAGATGTGAGAAACA 480  
DB 620 AGTTCAATTTCTGTGTTGCGGAGACACTGCACTGCTCCAGAGTTGAACCTTT 679  
DB 481 AGTTCAATTTCTGTGTTGCGGAGACACTGCACTGCTCCAGAGTTGAACCTTT 540  
DB 680 TTTGTTGTCAGAGTGTAGAGAGTCCCTTCCTCAATATATATGTTTGCATAGTTTC 729  
DB 541 TTTGTTGTCAGAGTGTAGAGAGTCCCTTCCTCAATATATATGTTTGCATAGTTTC 600  
DB 740 TTTTCAAGT 749  
DB 601 CTTTCAAT 610

RESULT 11  
LOCUS BQ421087 1125 bp mRNA linear EST 23-MAY-2002  
DEFINITION AGENCOURT\_7766236 NIH\_MGC\_68 Homo sapiens cDNA clone IMAGE:6021844  
5', mRNA sequence.  
ACCESSION BQ421087  
VERSION BQ421087.1 GI:21116402  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
REFERENCE 1 (bases 1 to 1125)  
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: cgapbs-remail.nih.gov  
Tissue Procurement: DCTD/DTF/Gazdar  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
DNA Sequencing by: Agencourt Bioscience Corporation  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LLNL at:  
http://image.llnl.gov  
Plate: LLM1328 row: d column: 05  
High quality sequence start: 284  
High quality sequence stop: 651.  
Location/Qualifiers  
1. 1125  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6021844"







Db 420 TCATCACCCTAGCTCTGTAGAGAAATGCAACCTCAGCTCTCAAGATGTGAGAACCA 479  
 Oy 618 CAAGTTCATTTCTGTGTGGGAGACACAGCTCAGCTCAGCTGAGGAGTGAAGTCT 677  
 Db 480 CAATTCATTTCTGTGTGGGAGACACAGCTCAGCTCAGCTGAGGAGTGAAGTCT 677  
 Oy 678 TTTTGTGTCAGAGTCTAGAGTCCCTTCTCTGATATATATCTGTTTGTCAATGTT 737  
 Db 539 TTTTGTGTTTCAAGTCTAGAGTCCCTTCTCTGATATATATCTGTTTGTCAATGTT 737  
 Oy 738 CTTTCAAGTCTAGAGTCCCTTCTCTGATATATATCTGTTTGTCAATGTT 767  
 Db 598 CTTTCAAGTCTAGAGTCCCTTCTCTGATATATATCTGTTTGTCAATGTT 627

RESULT 13  
 LOCUS BE794429 882 bp mRNA linear EST 20-SEP-2000  
 DEFINITION 601589445F1 NIH\_MGC\_7 Homo sapiens CDNA clone IMAGE:3943762 5'  
 mRNA sequence.  
 ACCESSION BE794429 GI:10215627  
 VERSION BE794429  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 882)  
 NHI-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: DCTD/DRP  
 CDNA Library Preparation: Ling Hong/Rubin Laboratory  
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL at: image.lnl.gov  
 Plate: LICM799 row: 1 column: 11  
 High quality sequence stop: 756.

FEATURES  
 source location/Qualifiers

1. 882  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3943762"  
 /clone.lib="NIH\_MGC\_7"  
 /tissue\_type="small cell carcinoma"  
 /cell\_line="MGC3"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: lung; Vector: pOT7; Site\_1: XhoI; Site\_2:  
 EcoRI; CDNA made by oligo-dr priming. Directionally  
 adaptor: GGCACAG(G). Size selected >500bp for average  
 insert size 1.8kb. Library constructed by Ling Hong in  
 the laboratory of Gerald M. Rubin (University of  
 California, Berkeley) using ZAP-cDNA synthesis kit  
 (Stratagene) and Superscript II RT (Life Technologies)."  
 BASE COUNT 190 a 246 c 237 g 209 t  
 ORIGIN

Query Match 71.3%; Score 591.8; DB 12; Length 882;  
 Best Local Similarity 97.0%; Pred. No. 2.2e-153;  
 Matches 668; Conservative 0; Mismatches 12; Indels 9; Gaps 6;  
 Oy 41 CCTTCAGACCCCTGTGTCCCA-GGCAGAGAAATACCTGTGAGAGCCCTCTTCAT 99  
 Db 137 CCGCCACCCCTGTGTCCCAAGGAGAGAAATACCTGTGAGAGCCCTCTTCAT 196  
 Oy 100 AGAACAGAGATGATCTGTGATTTCAAGACCTATGATCAGGCTCGAGAGCTGC 159  
 Db 197 AGAACAGAGATGATCTGTGATTTCAAGACCTATGATCAGGCTCGAGAGCTGC 256

Oy 160 TGAGAGTTTGTCAATGCTCTACTACACCATATGATAGGGGCGCTTGTCTCCG 219  
 Db 257 TGAGAGTTTGTCAATGCTCTACTACACCATATGATAGGGGCGCTTGTCTCCG 316  
 Oy 220 CCTGTACATGGGACAGCCACCTGTGTGGAATGCAATGCTGTTTGTGAGCAAGATC 279  
 Db 317 CCTGTACATGGGACAGCCACCTGTGTGGAATGCAATGCTGTTTGTGAGCAAGATC 376  
 Oy 280 CTTAGAGAGTTTGTGAAATGTTGCTTCCAGCAGTTTCAAAATCAGGCTGAGACTG 339  
 Db 377 CTTAGAGAGTTTGTGAAATGTTGCTTCCAGCAGTTTCAAAATCAGGCTGAGACTG 436  
 Oy 340 CCAGCTGTTTGTATGATGAGCCACACCAAGCAGCAGGCTCTTGTGTCTATCTGTG 399  
 Db 437 CCAGCTGTTTGTATGATGAGCCACACCAAGCAGCAGGCTCTTGTGTCTATCTGTG 496  
 Oy 400 ATCAGTGAAGTTTGTGAGGGAACAACAACGAGCTTCAACGAACTTCACTGACCGC 459  
 Db 497 ATCAGTGAAGTTTGTGAGGGAACAACAACGAGCTTCAACGAACTTCACTGACCGC 556  
 Oy 460 CCAAGCTTCAACCAACAGCAGTGTGGAAGATGCAAGTGTCTTCCGCTTCCAGGA 519  
 Db 557 CCAAGCTTCAACCAACAGCAGTGTGGAAGATGCAAGTGTCTTCCGCTTCCAGGA 616  
 Oy 520 CTGGGCGAGTATGAGGGG-TGGCAGAGTCTCTTCTTCTTCAATCAGCCTTGTGTA 578  
 Db 617 CTGGGCGAGTATGAGGGG-TGGCAGAGTCTCTTCTTCTTCAATCAGCCTTGTGTA 675  
 Oy 579 GAGAAATGCAACCTGACCTCTCAA-GAATGTGAGGAACAAGTATTTCTTGTG 637  
 Db 676 GAGAAATGCAACCTGACCTCTCAA-GAATGTGAGGAACAAGTATTTCTTGTG 734  
 Oy 638 CGAGAGACACTGACAGCTCAGCTGTCCGAGTGTGAATCTTTTGTGTCAAGTCTA 697  
 Db 735 CGAGAGACACTGACAGCTCAGCTGTCCGAGTGTGAATCTTTTGTGTCAAGTCTA 790  
 Oy 698 GAGTCCCTTCTCTGAAATATATCTGTT 726  
 Db 791 GGAGTCCCTTCTCTGAAATATATCTGTT 819

RESULT 14

BE463515/c 606 bp mRNA linear EST 27-JUL-2000  
 LOCUS BE463515  
 DEFINITION h24e08.x1 NCI-CGAP\_K1d1 Homo sapiens CDNA clone IMAGE:3183878 3'  
 similar to WP:719FA\_279.B CE22843 ; mRNA sequence.  
 ACCESSION BE463515  
 VERSION BE463515.1 GI:9509290  
 KEYWORDS EST.

SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 606)  
 NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.  
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),  
 Tumor Gene Index  
 Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgaabs-remail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LNL, send email to:  
 info@image.lnl.gov  
 Seq primer: -400p from Gibco  
 High quality sequence stop: 453.

FEATURES  
 source location/Qualifiers  
 1. 606

```

/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP_Kid11"
/lab_host="DH10B"
/notes="Organ: Kidney; Vector: p773p-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI-CGAP_Kid3 was
prepared, and 58 circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids 132376-132391, 1456007-145675, and
1500552-1502853). Subtraction by Bento Soares and M.
Fatima Bonalido."
BASE COUNT      171 a      144 c      150 g      141 t
ORIGIN
Query Match      71.2%; Score 590.6; DB 10; Length 606;
Best Local Similarity 98.5%; Pred. No. 4.3e-153;
Matches 596; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

167 TTGTGATGCTACTACACACCATGATAGCGGCGCTTGTCCCGCTGTAC 226
|||||
606 TTGTGATGCTACTACACCATGATAGCGGCGCTTGTCCCGCTGTAC 547
|||||
227 ATGGGACAGCCACCTGGTCTGGAATGCAATGCTGTTCAGACAAATCCTGAGT 286
|||||
546 ATGGGACAGCCACCTGGTCTGGAATGCAATGCTGTTCAGACAAATCCTGAGT 487
|||||
287 GAGTTTGAAGTGGCTTCACAGGAGTTCAAATCAGCGTGTGAGTGCAGGCT 346
|||||
486 GAGTTTGAAGTGGCTTCACAGGAGTTCAAATCAGCGTGTGAGTGCAGGCT 427
|||||
347 GTTCATGATGAAGCAGACACAGGAGTTCATCCTGTCATCTGATCAGT 406
|||||
426 GTTCATGATGAAGCAGACACAGGAGTTCATCCTGTCATCTGATCAGT 367
|||||
407 AAGTTGAGGGGAAACAACAGGAGTTCACAGGAGTTCATCCTGAGGAGGCT 466
|||||
366 AAGTTGAGGGGAAACAACAGGAGTTCACAGGAGTTCATCCTGAGGAGGCT 307
|||||
467 TCACCCAGCAACAGTGTGAAGATGCAAGTGCCTTCCTCCAGAGCTGGCC 526
|||||
306 TCACCCAGCAACAGTGTGAAGATGCAAGTGCCTTCCTCCAGAGCTGGCC 247
|||||
527 AGCTAGTGGGGGTGGCAGAGTCTCTTGTTCATTCAGCCCTGCTGTAGAGAA 586
|||||
246 AGCTAGTGGGGGTGGCAGAGTCTCTTGTTCATTCAGCCCTGCTGTAGAGAA 187
|||||
587 CAACCTGCACTCTCAAGATGTGAGAAACAAAGTCAATCTGTTGGGAGAGAC 646
|||||
186 CAACCTGCACTCTCAAGATGTGAGAAACAAAGTCAATCTGTTGGGAGAGAC 127
|||||
647 TGCAGACTCACTGTGCGGAGGTGAACCTCTTTTGTGTCAGAGTTCAGAGT 706
|||||
126 TGCAGACTCACTGTGCGGAGGTGAACCTCTTTTGTGTCAGAGTTCAGAGT 67
|||||
707 TTCCGATATATATCTGTTGTCTCATAGTTTCCCTTTCAAGTAACTTTATAT 766
|||||
66 TTCCGATATATATCTGTTGTCTCATAGTTTCCCTTTCAAGTAACTTTATAT 7
|||||
767 TTCTA 771
|||||
6 TTCTA 2

```

RESULT 15  
 BE616032  
 LOCUS  
 DEFINITION 601279506f1 NIH\_MGC\_39 Homo sapiens cDNA clone IMAGE:3611641 5',  
 mRNA sequence.  
 BE616032  
 BE616032

```

VERSION BE616032.1 GI:9897548
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS NIH-MGC http://mgi.nci.nih.gov/
TITLE 1 (bases 1 to 596)
JOURNAL National Institutes of Health, Mammalian Gene Collection (MGC)
COMMENT Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Ling Hong/Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.lnl.gov
Plate: LCM269 row: k column: 02
High quality sequence stop: 589.
FEATURES
Location/Qualifiers
1..596
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="NCI-CGAP_Kid11"
/lab_host="DH10B"
/notes="Organ: Kidney; Vector: p773p-Pac (Pharmacia) with
a modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI-CGAP_Kid3 was
prepared, and 58 circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneids 132376-132391, 1456007-145675, and
1500552-1502853). Subtraction by Bento Soares and M.
Fatima Bonalido."
BASE COUNT      137 a      150 c      154 g      154 t      1 others
ORIGIN
Query Match      71.0%; Score 589.4; DB 10; Length 596;
Best Local Similarity 99.7%; Pred. No. 9.3e-153;
Matches 590; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

137 GTGATCAGGCTGCAAGAGTCTGAGAGTGTTCATGATGTAAGTGTGCTTCA 196
|||||
1 GTGATCAGGCTGCAAGAGTCTGAGAGTGTTCATGATGTAAGTGTGCTTCA 60
|||||
197 AAGCGGGCGGCTTGTCTGTCGCGGCTGTACATGAGGAGCAGCAGCCTGCTGA 256
|||||
61 AAGCGGGCGGCTTGTCTGTCGCGGCTGTACATGAGGAGCAGCAGCCTGCTGA 120
|||||
257 AATGCTGTTTCAGCAGCAAGATCTGTAAGTGTGTTTGAATGTTGCTTCAAG 316
|||||
121 AATGCTGTTTCAGCAGCAAGATCTGTAAGTGTGTTTGAATGTTGCTTCAAG 180
|||||
317 TTCCAATCAGCGGTGACAGTCCAGCCTGTTTCATGATGTAAGCAGCAGCAG 376
|||||
181 TTCCAATCAGCGGTGACAGTCCAGCCTGTTTCATGATGTAAGCAGCAGCAG 240
|||||
377 AAGGCTCTGTTGTGATCATCTGATGAGTGTGAGAGGAGCAACAAAGGGAGCT 436
|||||
241 AAGGCTCTGTTGTGATCATCTGATGAGTGTGAGAGGAGCAACAAAGGGAGCT 300
|||||
437 AACCAAGACTTATCTGACCGCCAGCGCTCAGCAGCAGCAGCAGCAGCAGCAG 496
|||||
301 AACCAAGACTTATCTGACCGCCAGCGCTCAGCAGCAGCAGCAGCAGCAGCAG 360
|||||
497 AGTACTGCTTCCGCTTCCAGAGCTGGGAGGAGTGTGAGAGGAGTGTGAGAG 556
|||||
361 AGTACTGCTTCCGCTTCCAGAGCTGGGAGGAGTGTGAGAGGAGTGTGAGAG 420
|||||
557 TTCATCAGCCCTAGCTGTGAGAGAAAGCAAACTGACTCTCAAGAGATGTGAGAG 616
|||||

```

```

Db      421 |||||
          TTCATTCACCCCTAGCTGTGTAGAGAAATGCAAACTCGACTCTCAAGGATGTGAGAAC 480
Oy      617 ACAAATTCATTTCTGTGTGTGGGAGACACTGCAGACTCCACTGTGCCGAGGTTGAATC 676
          |||||
Db      481 ACAAATTCATTTCTGTGTGTGGGAGACACTGCAGACTCCACTGTGCCGAGGTTGAATC 540
          |||||
Oy      677 TTTTGTGCTCAGCTCTAGAGACTCCCTTTCTGAAATATATACTTGTG 728
          |||||
Db      541 TTTTGTGCTCAGCTCTAGAGACTCCCTTTCTGAAATATATACTTGTG 592
          |||||

```

Search completed: June 28, 2003, 03:34:20  
 Job time: 1399.27 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 23:39:51 ; Search time 213.801 Seconds  
(without alignments)

8742.532 Million cell updates/sec

Title: US-09-763-902b-10

Perfect score: 830  
Sequence: 1 tgaacttgccatccatg.....ataattaataacacatg 830

Scoring table: IDENTITY\_NOC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 seqs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0

Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq\_101002.\*

1: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.\*  
24: /SID2/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	830	100.0	830	21	AAA08036 Human protein tran
2	761	91.7	864	22	AAS26215 Human CDNA encodin
3	317	38.2	318	23	AAS58525 Human CDNA #1201 encodin
4	297.4	35.8	359	16	AAT22887 Human gene signatu
5	256.2	30.9	716	22	AAH07014 Human CDNA clone (
6	251.6	30.3	1146	21	AAA08041 Human protein tran
7	251.6	30.3	2581	24	ABA05754 Human NTP associa
8	251.6	30.3	2692	22	AAH17938 Human CDNA sequenc
9	251.6	30.3	2994	23	ABV24535 Human prostate exp

10	250	30.1	1387	22	AAS26214 Human CDNA encodin
11	223	26.9	247	24	ABL62553 Colon adenocarcino
12	132.6	16.0	133	19	AAH10492 Human biallelic po
13	113	13.6	8918	22	AAS26708 Human genomic DNA
14	113	13.6	8919	22	AAS26709 Human genomic DNA
15	105.2	12.7	1035	23	AAS75617 DNA encoding novel
16	76.4	9.2	805	23	ABL16401 Drosophila melanog
17	60	7.2	60	24	ABN38042 Human spliced tran
18	58.8	7.1	2932	23	ABL16400 Drosophila melanog
19	58.8	7.1	4050	23	ABL09616 Drosophila melanog
20	58.8	7.1	7883	23	ABL08308 Drosophila melanog
21	52.8	6.0	65	24	ABN53496 Mouse spliced tran
22	49.4	6.0	51	22	AAI27777 Human SNP oligonuc
23	39	4.7	50	22	AAI34644 Human SNP oligonuc
24	38	4.6	753	21	AAI39195 Zea mays DNA fragm
25	37.2	4.5	16633	24	ABL79985 Human chemically m
26	37	4.5	4993	24	ABL34103 Human immune syste
27	36.2	4.4	1050	22	AAH41175 Mammalian Cre reco
28	36.2	4.4	6117	24	ABL33024 Human immune syste
29	36	4.3	285	21	AAH00696 Human colon cancer
30	36	4.3	9666	24	ABL70256 Human immune syste
31	36	4.3	9666	24	ABL33085 Human metastasis a
32	36	4.3	9666	24	ABL34531 Human metastasis a
33	35.8	4.3	6113	24	ABN79989 Human prostate exp
34	35.8	4.3	15518	24	ABL70607 Chemically treated
35	35.8	4.3	15518	24	ABL34172 Human immune syste
36	35.8	4.3	15518	24	ABL34624 Human metastasis a
37	35.8	4.3	32191	22	AAI37009 Human musculoskele
38	35.4	4.3	506	23	ABV56923 Human prostate exp
39	35.4	4.3	4415	24	ABL70423 Chemically treated
40	35.4	4.3	4415	24	ABL33906 Human immune syste
41	35.4	4.3	4415	24	ABL34602 Human metastasis a
42	35.2	4.2	2149	22	AAK94875 Human full-length
43	35.2	4.2	116624	19	AAV52850 Human eyal gene co
44	35.2	4.2	151826	21	AAV22281 BAC containing rep
45	35	4.2	3156	24	ABL57129 Bacillus subtilis

## ALIGNMENTS

RESULT 1	AAA08036	standard; CDNA: 830 BP.
ID	AAA08036;	
AC	AAA08036;	
XX		
DT	19-JUN-2000	(first entry)
DE	Human protein transport molecule (PTM) encoding CDNA SEQ ID NO:10.	
XX		
KW	Human: protein transport molecule; PTM: diagnosis; cytosolic;	
KW	antiarthritic; antidiabetic; immunosuppressant; antileukocytic;	
KW	antiallergic; antidiabetic; antileukemic; antileukemic; osteoporotic;	
KW	dermatological; antidiabetic; antileukemic; antileukemic; antileukemic;	
KW	antileukemic; antileukemic; antileukemic; antileukemic; antileukemic;	
KW	immune disorder; cell proliferative disorder; secretory disorder;	
KW	uticaria; allergy; abnormal vesicle trafficking; asthma;	
KW	autoimmune haemolytic anaemia; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	WO200012703-A2.	
XX		
PD	09-MAR-2000.	
XX		
PF	26-AUG-1999; 99WO-US19616.	
XX		
PR	27-AUG-1998; 98US-0098206.	
XX		
PA	(INCY-) INCYTE PHARM INC.	
XX		
PI	Tang YT, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ;	





CC prevent skin aging due to sunburn, to maintain organs before  
 CC transplantation, for supporting cell culture of primary tissues, to  
 CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
 CC as a food additive or preservative to increase or decrease storage  
 CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
 CC minerals, cofactors and other nutritional components. The present  
 CC sequence encodes a novel secreted protein of the invention.

Query Match 91.7%; Score 761; DB 22; Length 864;  
 Best Local Similarity 100.0%; Pred. No. 1e-226;  
 Matches 761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 70 GGAATACCTGGTGGAGCCCTCTTCATAGAACAGAGATGCAATCTGTGATTCA 129
DB 35 GGAATACCTGGTGGAGCCCTCTTCATAGAACAGAGATGCAATCTGTGATTCA 94
QY 130 GACCTATGAGTGCAGAGCTGAGAGCTGAGAGTTTTCATATGCTACTACACAC 189
DB 95 GACCTATGAGTGCAGAGCTGAGAGCTGAGAGTTTTCATATGCTACTACACAC 154
QY 190 CATGATTAAGCGGCGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 249
DB 155 CATGATTAAGCGGCGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 214
QY 250 GAATGCAATGCTGTTTTCAGACAGAAATCCTGAGTGAAGTTTTCATATGCTGCT 309
DB 215 GAATGCAATGCTGTTTTCAGACAGAAATCCTGAGTGAAGTTTTCATATGCTGCT 274
QY 310 CAGCGATTCACAAATCAGAGCTGTGATGACGACGCTGCTGCTGCTGCTGCTGCT 369
DB 275 CAGCGATTCACAAATCAGAGCTGTGATGACGACGCTGCTGCTGCTGCTGCTGCT 334
QY 370 CCAGACACAGGTCCTGTTGTCATCTGTGATCAGTGAAGTTTTCATATGCTGCTG 429
DB 335 CCAGACACAGGTCCTGTTGTCATCTGTGATCAGTGAAGTTTTCATATGCTGCTG 394
QY 430 GGACTTCAACCAAGTCTATCTGACGCGCCAGGCTTCCAGCAGACAGTGTGAA 489
DB 395 GGACTTCAACCAAGTCTATCTGACGCGCCAGGCTTCCAGCAGACAGTGTGAA 454
QY 490 GATCGCAAGTGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 549
DB 455 GATCGCAAGTGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 514
QY 550 TCTTGTCTTCAATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 609
DB 515 TCTTGTCTTCAATGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 574
QY 610 GAGGAAACAGTATTTTGTGTTGGGAGACAGTGCAGACTGCTGCTGCTGCTG 669
DB 575 GAGGAAACAGTATTTTGTGTTGGGAGACAGTGCAGACTGCTGCTGCTGCTGCT 634
QY 670 TGAATCTTTTGTGCTCAAGTCTGAGAGTCCCTTCCGATATATATCTTGTGT 729
DB 635 TGAATCTTTTGTGCTCAAGTCTGAGAGTCCCTTCCGATATATATCTTGTGT 694
QY 730 CATAGTCTTCTTCAAGTGAAGTATTTTCTATTTTCTACTGCTGCTGCTGCTG 789
DB 695 CATAGTCTTCTTCAAGTGAAGTATTTTCTATTTTCTACTGCTGCTGCTGCTG 754
QY 790 TGATTCGTGAATCTGCAAAATATATATATATATATATATATATATATATAT 830
DB 755 TGATTCGTGAATCTGCAAAATATATATATATATATATATATATATATATAT

```

RESULT 3  
 AAS58525 standard; cDNA: 318 BP.  
 AAS58525;  
 13-FEB-2002 (first entry)

DE cDNA #1201 encoding portion of a human colon tumour protein.

XX Human: colon tumour protein; colon cancer; gene therapy; cytostatic; ss.  
 KW Homo sapiens.  
 OS Homo sapiens.  
 PN MO200173027-A2.

PD 04-OCT-2001.

PF 22-MAR-2001; 2001MO-US09246.

PR 24-MAR-2000; 2000US-191597P.

PR 04-MAY-2000; 2000US-202024P.

PR 05-MAY-2000; 2000US-202189P.

(CORI-) CORIXA CORP.

Meagher MJ, Xu J, King GE;

WPI; 2001-611627/70.

New colon tumour proteins and related nucleic acid, useful for

treatment, prevention, diagnosis and monitoring of cancer.

Claim 4; Page 246; 299pp; English.

Th present invention relates to the isolation of novel cDNA sequences  
 encoding for at least an immunogenic portion of human colon tumour  
 proteins. The sequences of the invention are useful in pharmaceutical  
 compositions and vaccines for the prevention and treatment of cancers  
 such as colon cancer. They are also useful for the diagnosis and  
 monitoring of such cancers. Antibodies to the colon tumour proteins  
 and antigen presenting cells that express polynucleotides encoding  
 colon tumour proteins can be used to inhibit the development of  
 cancers. T-cells that react specifically with colon tumour proteins  
 are useful for removing tumour cells from samples (e.g. blood) and  
 for cancer treatment. The polynucleotide sequences are also useful in  
 gene therapy. AAS7325-AAS5880 represent the cDNA sequences of the  
 invention that encode for portions of human colon tumour proteins.

Sequence 318 BP; 76 A; 82 C; 83 G; 76 T; 1 other;

Query Match 38.2%; Score 317; DB 23; Length 318;  
 Best Local Similarity 99.7%; Pred. No. 2e-88;  
 Matches 317; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

QY 147 CCTGAGAGCTGCTGAGAGTGTGCAATGCTACTACACCACTGATAGCGGCGC 206
DB 1 CCGTGAAGCTGCTGAGAGTGTGCAATGCTACTACACCACTGATAGCGGCGC 60
QY 207 GTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 266
DB 61 GTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 120
QY 267 CAGGACAGAAATCCTGAGTGAAGTATTTTGAATGTTGCTTCCAGCAATCA 326
DB 121 CAGGACAGAAATCCTGAGTGAAGTATTTTGAATGTTGCTTCCAGCAATCA 180
QY 327 GCGTGTGAGTGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 386
DB 181 GCGTGTGAGTGCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
QY 387 TTGTCACTGTGATCAATGATGATGATGATGATGATGATGATGATGATGATG 446
DB 241 TTGTCACTGTGATCAATGATGATGATGATGATGATGATGATGATGATGATG 300
QY 447 TCATCTGAGCGGCCAGG 464
DB 301 TCATCTGAGCGGCCAGG 318

```

RESULT 4





XX	Sequence	716 BP; 231 A; 148 C; 171 G; 163 T; 3 other:
XX	Query Match	30.9%; Score 256.2; DB 22; Length 716;
XX	Best Local Similarity	74.3%; Pred. No. 2.9e-69;
XX	Matches 321; Conservative	0; Mismatches 111; Indels 0; Gaps 0
OY	114 CATCTGTGATTTTCAGACATCTATGATTCAGGACCTGCGAGACCTGCTGAGAGTTTGCA	173
DB	275 CGTCTTGATTTTAAACTTATGTAGATCAGGCAATGTAGACTGCTGAGAGTTTGCA	334
OY	174 ATGCTACTACACCCATGATTAAGCGCGGCGGCTTTGCTGCTCCGCTGTACATAGGCA	233
DB	335 ATATTACTATGAGACATATGATTAAGAAGACGGGCACTAACAGGCTGTATCTGGACA	394
OY	234 CAGCCACCTGCTCTGTAATGGCAATGCTGTTCACAGACAAAGATCTTGAGTGAATT	293
DB	395 AGGCCACCTTATATGAAATGAAATGCTGTTCAGGGGTGATGCCCTTAATTAATTTT	454
OY	294 TTGAATTTGGCTTCCAGCGAGTTCCAATACAGCTGGTAGACTCCAGCTGTTCATG	353
DB	455 TTGACACATGCTCTTCTAGTAGAGTTCCAGCTCAATATGTATAGTTGCCAACAGCTTCATG	514
OY	354 ATGAAGCCACACCAAGCCAGACAGCGTCTGTTGTCTATCTGTGTGATCAGTGAAGTTG	413
DB	515 ACCAAGCACTAGTCCCAACTAGAGTTCTTGTGTGACCAAGTGAAGCTGTAAATTTG	574
OY	414 AGGGGAACAACACGGGACTTCACCAAGACTTCAATCTGACCGCCAGGCTCACCCA	473
DB	575 ATGGAACAACAACAACCTTCTTCACACAGAACTCTCTGCTGACTGCTCACTCCCA	634
OY	474 GCAACACATGTGGGAAGATCGCAAGTACGCTTCGCTTCAGAGCTGGCCAGCTAGT	533
DB	635 ACAATATCTGTGGAAATTTGCAAGTATGCTTCCGTTTCAAGAAATGGCCAAATANT	694
OY	534 GGGGGTGCAGA 545	
DB	695 TTPANGGCCAAA 706	
RESULT 6		
ID	AAA08041	standard; cDNA; 1146 BP.
AC	AAA08041;	
DT	19-JUN-2000	(first entry)
DE	Human protein transport molecule (PTM) encoding cDNA seq ID NO:15.	
XX	Human; protein transport molecule; PTM; diagnosis; cytostatic;	
KW	antiatheritic; antiasthmatic; immunosuppressant; antiarteriosclerotic;	
KW	antiallergic; antidiabetic; antilipemic; antirheumatic; osteopathic;	
KW	dermatological; antihiv; protein transport regulator; cancer;	
KW	antitumorigenic; antihiv; protein transport regulator; cancer;	
KW	immune disorder; cell proliferative disorder; secretory disorder;	
KW	urticaria; allergy; abnormal vesicle trafficking; asthma;	
XX	autoimmune haemolytic anaemia; ss.	
OS	Homo sapiens.	
PN	WO200012703-A2.	
PD	09-MAR-2000.	
PF	26-AUG-1999; 99WO-US19616.	
PR	27-AUG-1998; 9805-0098206.	
PA	(INCY-) INCYTE PHARM INC.	
PI	Tang YT, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ;	
PI	Gorgone GA, Baughn MR, Patterson C.	

XX MPI: 2000-256642/22.  
dr P-PSDB; AAY82323.

XX New human protein transport-associated polypeptide and polynucleotide  
PT useful for diagnosis, prevention and treatment of cell proliferative  
pr and secretory disorders such as leukemia, cystic fibrosis

PS Claim 9; Page 74; 75pp; English.

XX AAA08035 to AAA08042 encode the human protein transport-associated  
CC molecules (PRAMs) given in AY82317 to AY82332. The PRAMs have  
CC cytosolic, antithratic, antiaesthetic, immunosuppressant,  
CC antileukosteric, antiallergic, antidiabetic, antilipemic,  
CC antirheumatic, osteopathic, dermatological and anti-HIV activities, and  
CC hepatotropic, antigenic, antiflammatory and antitumor activities, and  
CC regulate protein transport. PRAM proteins and antagonists are useful for  
CC preventing or treating a disorder associated with decreased or increased  
CC expression or activity of PRAM. PRAM polynucleotides are useful for  
CC diagnosing conditions associated with PRAM, comprising detecting PRAM by  
CC forming a hybridisation complex, preferably after PCR amplifying the  
CC biological sample. Diseases prevented, treated or diagnosed include cell  
CC proliferative disorders such as cancers, immune disorders, secretory  
CC disorders and other conditions associated with abnormal vesicle  
CC trafficking, such as allergies, asthma, urticaria and autoimmune  
CC haemolytic anemia. Anti-PRAM antibodies may be used as antagonists, as  
CC a targeting or delivery mechanism for bringing pharmaceutical agents  
CC into cells or tissues expressing PRAM and for diagnosis of PRAM-related  
CC disorders. PRAM, its catalytic or immunogenic fragments are useful for  
CC drug screening using libraries of compounds. PRAM polynucleotides are  
CC useful for generating hybridisation probes useful in mapping the  
CC naturally occurring genomic sequences.

XX Sequence 1146 BP: 331 A: 212 C: 228 G: 375 T: 0 other:

SQ

Query Match 30.3%; Score 251.6; DB 21; Length 1146;  
Best Local Similarity 72.4%; Pred. No. 1e-67;  
Matches 326; Conservative 0; Mismatches 124; Indels 0; Gaps 0

OY 114 CATCTGTGATTTTCAGACCTATGTGGATCAGGCGCTGCAGAGCTGTGAAGATTGTCA 173  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
162 CGCTGTGGATTTTTAAACTTATGTAAGTAGCATGTAGAGTGCTGAGAGATTGTCA 221  
OY 174 ATGTCTACTACACACCACCATGTAGATAAGCGGCGGCTTGTGCCCGCTGTACATGGCA 233  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
222 ATTATTACTRTGAGACAATGGATGAAAAAAGAAGGGCACATAACACAGCTGTATCTGACA 281  
OY 234 CAGCCACCCCTGTCTGTGGAATGCAATGCTGTTTCAGACACAGAAATCTTGAGTCAATT 293  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
282 AGGCCACCTTAATATGGAATGAAAATGCTGTTTCAGGCGCTGAGTGCCTTAATAATTTT 341  
OY 294 TTGAATGTGCTCTCCAGGAGATTCCAATCAGCGGTGATGAGATGCGACCTGTTCATG 353  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
342 TTACACATTGCTTTAGTAGATCTCCAGGTCAATATGTTAGTTGCAACACAGTTTCATG 401  
OY 354 ATGAGCAGACACCAAGCAGACAGCGGCTTGTGTCATCTGTGATCAGTGAAGTTTG 413  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
402 AGCAAGCACTCAGTCCCAAACACTACATTTGTTGTGACACAGTGAAGTGAAGTTTG 461  
OY 414 AGGGGAGCAACAACAGGACTTCAACCAAGTTCATCTGACCGCCAGGCTCACCA 473  
DB | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |  
462 ATGGAACAACAACAATTTCTTCACACAGAACTCCGTGAGTGTGCTCACTCACATCCCA 521  
OY 474 GCAACACAGTGTGAAGATGCGCAGTACAGTCTCCGCTCCAGGATGGGCGAGTAGT 533  
DB 522 ACATATCTGTGTGAAATATGCAAGTATGCTCCGTTTTCANAGATTGGTCTAGTAGTT 581  
OY 534 GGGGGTGACAGAGTCTCTTGTGCTTCAATC 563  
DB 582 AAAGGGCAAAAGTCCATCTCATTTGGTC 611

RESULT 7

ABA05754  
 ID ABA05754 standard; cDNA; 2581 BP.  
 XX  
 AC ABA05754;  
 XX  
 DT 04-MAR-2002 (first entry)  
 XX  
 DE Human NTF2 associated protein 16 coding sequence.  
 XX  
 KW Human; NTF2 associated protein 16; cancer; HIV infection; gene therapy;  
 KM ss.  
 XX  
 OS Homo sapiens.  
 XX  
 FH Key Location/Qualifiers  
 FT CDS 147..575  
 FT /tag-a  
 FT /product- "NTF2 associated protein 16"  
 XX  
 PN CNI33332-A.  
 XX  
 PD 19-SEP-2001.  
 XX  
 PF 15-MAR-2000; 2000CN-0114918.  
 XX  
 PR 15-MAR-2000; 2000CN-0114918.  
 XX  
 PA (BODE-) BODE GENE DEV CO LTD SHANGHAI.  
 XX  
 PI Mao Y, Xie Y;  
 XX  
 DR WPI: 2002-042199/06.  
 DR P-PSDB; ABB04459.  
 XX  
 PT Polypeptide-human NTF2 associated protein 16 and polynucleotide  
 PT encoding it -  
 XX  
 PS Claim 6; Page 24-25(Disclosure); 32pp; Chinese.  
 CC The present invention provides the protein and coding sequences of human  
 CC NTF2 associated protein 16. The sequences can be used in the treatment of  
 CC cancer and HIV infection, as well as other diseases. The present sequence  
 CC is the coding sequence of the invention.  
 XX  
 SQ Sequence 2581 BP; 791 A; 449 C; 452 G; 889 T; 0 other;  
 XX  
 Query Match 30.3%; Score 251.6; DB 24; Length 2581;  
 Best Local Similarity 72.4%; Pred. No. 1.7e-67;  
 Matches 326; Conservative 0; Mismatches 124; Indels 0; Gaps 0;  
 XX  
 OY 114 CATCTGATTTCAAGACCTATGTGATCAGGCTGCGAGAGCTGAGAGTTTGTCA 173  
 DB 154 CGCTCTGATTTTAAACTATGTAGATCAGCAGATGATGCTGAGAGTTTGTCA 213  
 OY 174 AGTCTACTACACACCATGATGATGAGGCGGCTTGTGCTCCCGCTGTACATGGCA 233  
 DB 214 ATATTACTATGAGCAATGATGATAAAGAGAGCGGCGACTACACAGCTGTATCTGACA 273  
 OY 234 CAGCCACCCGTGCTGTGAATGCAATGCTGTTTCAAGACAGATCTTGTAGTGAATTTT 293  
 DB 274 AGGCCACCTTAATATGAAATGAAATGCTGTTTCAAGGCGTGGATGCCCTAAATATTTT 333  
 OY 294 TTGAATGTTGCTTCCAGCGAGATTCGAATCAGCGTGTGAGACTGCCAGCTGTTTCA 353  
 DB 334 TTGACACATTTGCTTGTAGAGATTCAGGTCAATATGTTAGATTGCCAACAGTTTCAT 393  
 OY 354 ATGAACCCACACACACACACAGCGCTGTTGTGCTGATCTGTGATGATGAATTTG 413  
 DB 394 AGCAACCACTAGTCCCAACACTACAGTCTTGTGTTGACCACTGCAACTGTAAGTTTG 453  
 OY 414 AGGGGAAACAACAGGAGCTTCAACAGAACTTCACTGAGCGGCCAGGCTCAACCCA 473  
 DB 454 ATGGAACAAACAACATTTCTTCAACAGAACTTCTGCTGAGTGTCACTGCCATGCCA 513

OY 474 GCAACACAGTGTGAGAGATGCAAGTACGCTTCGCTTCAGAGACTGGCCAGCTACT 533  
 DB 514 ACAATAGTGTGTGAGATGCAAGTATGCTTCGCTTTCAGAGATGCTAGTACTT 573  
 OY 534 GGGGTGGCAGAGTCTTTCCTTTCATTC 563  
 DB 574 AAGGGGCAAAAGTCTTCATTCATTGCTC 603  
 XX  
 RESULT 8  
 ID AAH17938  
 XX AAH17938 standard; cDNA; 2692 BP.  
 AC AAH17938;  
 XX  
 DT 26-JUN-2001 (first entry)  
 XX  
 DE Human cDNA sequence SEQ ID NO:17706.  
 XX  
 KW Human; primer; detection; diagnosis; antisense therapy; gene therapy; ss.  
 XX  
 OS Homo sapiens.  
 XX  
 PN EP1074617-A2.  
 XX  
 PD 07-FEB-2001.  
 XX  
 PF 28-JUL-2000; 2000EP-0116126.  
 XX  
 PR 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 XX  
 DR WPI: 2001-318749/34.  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs -  
 XX  
 PS Claim 8; SEQ ID 17706; 2537pp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dt primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH1633 to AAH18742 represent human cDNA sequences; AAB92446 to  
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX



KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW vulnery; secreted protein; rheumatoid arthritis;  
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KW cerebrovascular disorder; cerebral ischemia; angiodogenesis;  
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KW corneal infection; wound healing; epithelial cell proliferation;  
KW skin ageing; food additive; preservative; antiproliferative.  
XX  
XX Homo sapiens.  
PN WO200155322-A2.  
XX  
XX  
PD 02-AUG-2001.  
PF 17-JAN-2001; 2001WO-US01341.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226686.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.

PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0234999.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239335.  
PR 13-OCT-2000; 2000US-0239337.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241809.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249225.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.

Mon Jun 30 08:51:07 2003

PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 03-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
PI WPI; 2001-488783/53.  
DR P-PSDB; AAU16227.  
XX  
XX New nucleic acid molecules encoding 461 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives -  
XX  
XX Claim 1; SEQ ID NO 393; 980pp; English.  
XX  
XX The invention relates to isolated nucleic acid molecules and their  
CC encoded secreted proteins. The nucleic acids and proteins are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition of susceptibility  
CC to a pathological condition. Antibodies to the proteins can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbant assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischaemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection, and many other  
CC disorders listed in the specification. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence encodes a novel secreted protein of the invention.

Query Match 30.1%; Score 250; DB 22; Length 1387;  
Best Local Similarity 71.0%; Pred. No. 3.7e-67;  
Matches 331; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

QY 98 ATAGAACACAGATGGCATCTGTGGATTTCAGACCTATGGATCAGCGCTGACAGCT 157  
DB 52 ATAGCGAGAGCGGGGGTGTGGGGATTTAAACTTATGTAGATCAGCGATGTAGACT 111  
QY 158 GCTGAGGAGTTTGTCAATGTCTACTACACCATGGATTAAGCGGCGGTTTGTCTGCC 217  
DB 112 GCTGAGGAGTTTGTCAATGTCTACTATGACAAATGGATTAAGAGAGCGGCACTAAC 171  
QY 218 GCGCTGTATACATGGGACAGCCACCCCTGGTCTGGAATGGCAATGCTGTTTCAGGACAA 277  
DB 172 AGGCTGTATCTGCAGAACGCCACCTTAATATGGAATGGAATGCTGTTTCAGGCGTGGAT 231  
QY 278 TCGTTGAGTGAATTTTGAATGTGCTTCCAGCGAGTTCCCAATCAGCGTGTAGAC 337  
DB 232 GCGCTAAATTAATTTTGTGACACATGCGCTTCTAGTGAGTTCCAGTCAATGTAGAT 291  
QY 338 TCCGACGCTGTTCATGATGAAGCCACCAAGCCAGCCAGCCAGCTTCTTGTCTCATCTGT 397  
DB 292 TCCCAACAGATTCATGAGCAAGCACTAGTCCCAACTACAGTTCTTGTGTGACCACT 351  
QY 398 GGATCAGTGAAGTTTGTAGGGGGAACAAACACGGGACTTCAACAGAACTTCATCTGACC 457  
DB 352 GGAAGTGTGAAGTTTGTAGGGAACAAACAACTTTCTTCAACAGCAAGTTCCTGCTGACT 411

QY 458 GCCAGGCTCACCACCAACACAGTGTGGAAGATCGCAAGTCACTGCTTCCGCTTCCAG 517  
DB 412 GCTCAGTCCACCTCCCAACAACTACTGTGTGAAGATTGCAAGTATGCTTCCGCTTCCAA 471  
QY 518 GACTGGCCAGCTAGTGGGGTGGCAGAGGTCTCTTGTCTTCATTC 563  
DB 472 GATTGCTCTAGTAGTTAAAGGGGCAAAAGTCCATTCTCTATTGGTC 517  
RESULT 11  
ABL62553/C  
ID ABL62553 standard; DNA; 247 BP.  
XX  
XX ABL62553;  
XX  
XX 15-MAY-2002. (first entry)  
DT  
XX  
DE Colon adenocarcinoma related gene sequence SEQ ID NO:890.  
XX  
XX Human; cancer; colon; breast; ovary; oesophagus; kidney; thyroid;  
KW stomach; lung; prostate; pancreas; carcinoma; antitumour; cancerous;  
KW cytostatic; gene therapy; antineoplastic; Wilm's tumour; adenocarcinoma;  
KW gene; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO200194629-A2.  
PD  
XX 13-DEC-2001.  
XX  
XX 30-MAY-2001; 2001WO-US10838.  
XX  
XX 05-JUN-2000; 2000US-209473P.  
PR 05-JUN-2000; 2000US-209531P.  
PR 18-SEP-2000; 2000US-233133P.  
PR 18-SEP-2000; 2000US-233617P.  
PR 20-SEP-2000; 2000US-234009P.  
PR 20-SEP-2000; 2000US-234034P.  
PR 20-SEP-2000; 2000US-234052P.  
PR 22-SEP-2000; 2000US-234509P.  
PR 22-SEP-2000; 2000US-234567P.  
PR 25-SEP-2000; 2000US-234923P.  
PR 25-SEP-2000; 2000US-234924P.  
PR 25-SEP-2000; 2000US-235077P.  
PR 25-SEP-2000; 2000US-235082P.  
PR 25-SEP-2000; 2000US-235134P.  
PR 25-SEP-2000; 2000US-235280P.  
PR 26-SEP-2000; 2000US-235637P.  
PR 26-SEP-2000; 2000US-235638P.  
PR 27-SEP-2000; 2000US-235711P.  
PR 27-SEP-2000; 2000US-235720P.  
PR 27-SEP-2000; 2000US-235840P.  
PR 27-SEP-2000; 2000US-235863P.  
PR 28-SEP-2000; 2000US-236028P.  
PR 28-SEP-2000; 2000US-236032P.  
PR 28-SEP-2000; 2000US-236033P.  
PR 28-SEP-2000; 2000US-236034P.  
PR 28-SEP-2000; 2000US-236109P.  
PR 28-SEP-2000; 2000US-236111P.  
PR 29-SEP-2000; 2000US-236842P.  
PR 29-SEP-2000; 2000US-236891P.  
PR 02-OCT-2000; 2000US-237172P.  
PR 02-OCT-2000; 2000US-237173P.  
PR 02-OCT-2000; 2000US-237278P.  
PR 02-OCT-2000; 2000US-237294P.  
PR 02-OCT-2000; 2000US-237295P.  
PR 02-OCT-2000; 2000US-237316P.  
PR 03-OCT-2000; 2000US-237425P.  
PR 03-OCT-2000; 2000US-237598P.  
PR 03-OCT-2000; 2000US-237604P.  
PR 03-OCT-2000; 2000US-237606P.  
PR 03-OCT-2000; 2000US-237608P.  
PR 01-NOV-2000; 2000US-244867P.

PR 01-NOV-2000; 200005-245084P.  
XX  
XX (AVAL-) AVALON PHARM.  
XX  
PI Young PE, Augustus M, Carter KC, Edner R, Endress G, Horrigan S;  
PI Soppet DR, Weaver Z;  
XX  
XX WPI: 2002-188264/24.  
XX  
PT Screening for anti-neoplastic agent involves exposing cells to a  
PT chemical agent to be tested for anti-neoplastic activity, and  
PT determining a change in expression of a gene of a signature gene set  
XX  
PS Claim 1; SEQ ID 890; 44pp; English.  
XX  
CC The present invention describes a method (M1) for screening for an  
CC anti-neoplastic agent. The method involves exposing cells to a chemical  
CC agent to be tested for anti-neoplastic activity, determining a change in  
CC expression of at least one gene (I) of a signature gene set, where (I)  
CC comprises a sequence (S) selected from 8447 sequences (given in ABL61664  
CC to ABL70110), or is at least 95% identical to (S), where a change in  
CC expression is indicative of anti-neoplastic activity. (I) has cytostatic  
CC activity and can be used in gene therapy. M1 can be used for screening  
CC an anti-neoplastic agent, and can be used for producing a product which  
CC is the data collected with respect to the anti-neoplastic agent as a  
CC result of M1, and the data is sufficient to convey the chemical  
CC structure and/or properties of the agent. M1 can be used in the  
CC treatment of cancer such as colon, breast, stomach, lung, thyroid,  
CC esophageal, ovarian, kidney, prostate or pancreatic cancer,  
CC adenocarcinoma, carcinoma, clear cell cancer, infiltrating ductal cancer,  
CC infiltrating lobular cancer, squamous cell carcinoma, neuroendocrine  
CC carcinoma, papillary carcinoma and Wilms' tumour.  
XX  
SQ Sequence 247 BP; 89 A; 41 C; 45 G; 72 T; 0 other;

Query Match 26.9%; Score 223; DB 24; Length 247;  
Best Local Similarity 100.0%; Pred. No. 3.6e-59;  
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 608 GTGAGACACACAAATTCATTTCTGTTGTCGGAGACACTGCAGACTCCACTGCCGAG 667  
DB 247 GTGAGACACACAAATTCATTTCTGTTGTCGGAGACACTGCAGACTCCACTGCCGAG 188  
OY 668 GTTGAAGCTCTTTTGTGCTCAAGTTCAGAGTCCCTCCGATATATATCTGTT 727  
DB 187 GTTGAAGCTCTTTTGTGCTCAAGTTCAGAGTCCCTCCGATATATATCTGTT 128  
OY 728 GTCATAGTTCCCTTTCAAGTAGTAACTTTCTATTTTCTACTTGCCAGTAGAGAC 787  
DB 127 GTCATAGTTCCCTTTCAAGTAGTAACTTTCTATTTTCTACTTGCCAGTAGAGAC 68  
OY 788 TCTGATTCGGAATTCGACAAATATTAATTAATACACATG 830  
DB 67 TCTGATTCGGAATTCGACAAATATTAATTAATACACATG 25

RESULT 12  
AAAX10492/c  
ID AAAX10492 standard; DNA; 133 BP.

AC AAAX10492;  
XX  
XX 30-MAR-1999 (first entry)  
XX  
DE Human biallelic polymorphic DNA fragment WI-13859.  
XX  
XX Polymorphism: biallelic; human; forensic; paternity testing; disease;  
KM detection; phenotypic typing; characteristic; infection; hereditary;  
KM autoimmune disease; cancer; inflammation; drug; therapy; medication;  
KM treatment; marker; ss.  
XX  
OS Homo sapiens.  
XX

PN MO9820165-A2.  
XX  
XX 14-MAY-1998.  
PD  
XX 05-NOV-1997; 97MO-US20313.  
PE  
XX 06-NOV-1996; 96US-0030455.  
PR  
XX  
XX (WHD) WHITEHEAD INST BIOMEDICAL RES.  
PA  
XX Hudson T, Lander ES, Wang D;  
PI  
XX  
XX WPI: 1998-286974/25.  
XX

PT New isolated nucleic acid segments from the human genome - used for  
PT determining polymorphic forms for use in e.g. forensics, paternity  
PT testing or phenotypic typing for disease  
XX  
PS Claim 1; Page 55; 310pp; English.

CC AAX10269-X12937 are human DNA fragments which contain biallelic  
CC polymorphic markers which have been isolated using the primers  
CC represented in AAX09121-X10268. The base occupying the polymorphic site  
CC is indicated by the appropriate IUPAC-IUB ambiguity code. These fragments  
CC can be used in methods for determining polymorphic forms in an individual  
CC for use in e.g. forensics, paternity testing or for phenotypic typing for  
CC diseases such as agammaglobulinemia, diabetes insipidus, Lesch-Nyhan  
CC syndrome, muscular dystrophy, Wiskott-Aldrich syndrome, Fabry's disease,  
CC familial hypercholesterolemia, polycystic kidney disease, hereditary  
CC spherocytosis, von Willebrand's disease, tuberous sclerosis, hereditary  
CC hemorrhagic telangiectasia, familial colonic polyposis, Ehlers-Danlos  
CC syndrome, osteogenesis imperfecta, acute intermittent porphyria,  
CC autoimmune diseases, inflammation, cancer, diseases of the nervous  
CC system, infection by pathogenic microorganisms, and characteristics such  
CC as longevity, appearance (e.g. baldness, obesity), strength, speed,  
CC endurance, fertility, and susceptibility or receptivity to particular  
CC drugs or therapeutic treatments. The isolated polymorphic nucleic acid  
CC segments can also be used to produce medicaments for the treatment or  
CC prophylaxis of such diseases.  
XX  
SQ Sequence 133 BP; 54 A; 20 C; 24 G; 34 T; 1 other;

Query Match 16.0%; Score 132.6; DB 19; Length 133;  
Best Local Similarity 99.2%; Pred. No. 4.1e-31;  
Matches 132; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 682 TGTGCTCAAGTCTAGAGTCCCTTCTCGAATATATATCTGTTGTCATAGTTCTT 741  
DB 133 TGTGCTCAAGTCTAGAGTCCCTTCTCGAATATATATCTGTTGTCATAGTTCTT 74  
OY 742 TTCAAAGTAACTTTCTATTTTCTACTTGCCAGTAGAGACTGATTCGAAA 801  
DB 73 TTCAAAGTAACTTTCTATTTTCTACTTGCCAGTAGAGACTGATTCGAAA 14  
OY 802 TTCTGACAAATAA 814  
DB 13 TTCTGACAAATAA 1

RESULT 13  
AAS26708/c  
ID AAS26708 standard; DNA; 8918 BP.

AC AAS26708;  
XX  
XX 07-NOV-2001 (first entry)  
XX  
DE Human genomic DNA encoding partial novel secreted protein, Seq ID 1682.  
XX  
XX Human; immunosuppressive; antiarthritic; ds; antirheumatic;  
KM cytosolic; cardiac; vasotropic; cerebroprotective; nootropic;  
KM neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KM vulnerable; secreted protein; rheumatoid arthritis;



KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KW cerebrovascular disorder; cerebral ischemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KW corneal infection; wound healing; epithelial cell proliferation;  
KW skin aging; food additive; preservative; antiproliferative.  
XX Homo sapiens.  
OS MO200155322-A2.  
FN 02-AUG-2001.  
PD 17-JAN-2001; 2001MO-US01341.  
XX 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226866.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 05-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 06-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231245.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231966.  
PR 14-SEP-2000; 2000US-0233397.  
PR 14-SEP-2000; 2000US-0233398.  
PR 14-SEP-2000; 2000US-0233399.  
PR 14-SEP-2000; 2000US-0233400.  
PR 14-SEP-2000; 2000US-0233401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.

PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234224.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 25-SEP-2000; 2000US-0234999.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 02-OCT-2000; 2000US-0237040.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246603.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251866.  
PR 08-DEC-2000; 2000US-0251869.







GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OW nucleic - nucleic search, using sw model

Run on: June 28, 2003, 00:24:04 ; Search time 1923.73 Seconds

(Without alignments)  
9647.960 Million cell updates/sec

Title: US-09-763-902B-15

Perfect score: 1146

Sequence: 1 tttacccggagagagagatggg.....tttaagtaattttgtttt 1146

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 16154066 seqs, 8097743376 residues

Total number of hits satisfying chosen parameters: 32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database : EST \*

1: em\_estba:\*  
2: em\_estchun:\*  
3: em\_estcin:\*  
4: em\_estcnu:\*  
5: em\_estcov:\*  
6: em\_estdpl:\*  
7: em\_estdro:\*  
8: em\_esthc:\*  
9: gb\_estl1:\*  
10: gb\_estl2:\*  
11: gb\_hic:\*  
12: gb\_est3:\*  
13: gb\_est4:\*  
14: gb\_est5:\*  
15: em\_estfun:\*  
16: em\_estgom:\*  
17: gb\_gss:\*  
18: em\_gss\_hum:\*  
19: em\_gss\_inv:\*  
20: em\_gss\_pln:\*  
21: em\_gss\_vrt:\*  
22: em\_gss\_fun:\*  
23: em\_gss\_mam:\*  
24: em\_gss\_mus:\*  
25: em\_gss\_other:\*  
26: em\_gss\_pro:\*  
27: em\_gss\_trod:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	847	73.9	934	9	AL539324
2	834	72.8	978	9	AL548271
3	824.4	71.9	993	14	BO051357
4	789.4	68.9	793	9	AL598693
5	775	67.6	881	11	AF212223
6	711.2	62.1	781	9	AU123481

7	706.2	61.6	770	9	AU137118	AU137118
8	682.8	59.6	745	12	BE887056	BE887056
9	681.4	58.5	782	13	B1458206	B1458206
10	666.8	58.2	721	10	AV114243	AV114243
11	656.4	57.3	743	13	B1761278	B1761278
12	649.2	56.6	819	13	B1597498	B1597498
13	612.2	53.4	878	12	BC035698	BC035698
14	611	53.3	629	12	BF087075	BF087075
15	611	53.3	629	13	BE995586	BE995586
16	610.4	53.3	713	10	AV118042	AV118042
17	587.4	51.3	884	12	BE205046	BE205046
18	579.4	50.6	750	12	BE718308	BE718308
19	562.4	49.1	666	10	BE390803	BE390803
20	560.8	48.9	805	12	BE204966	BE204966
21	558	48.7	702	13	B1828542	B1828542
22	551.6	48.1	996	12	BE480012	BE480012
23	541	47.2	931	12	BE717309	BE717309
24	526	45.9	526	14	BM833679	BM833679
25	523	45.6	1016	10	BE540322	BE540322
26	511	44.6	511	9	AA400543	AA400543
27	503.4	43.9	506	13	BM147972	BM147972
28	502.2	43.8	574	12	BE004693	BE004693
29	491.2	42.9	716	9	AU133525	AU133525
30	473.6	41.3	493	9	AL702205	AL702205
31	471	41.1	568	10	AV716990	AV716990
32	448.6	39.1	487	12	BF087098	BF087098
33	436	38.0	444	9	AA831754	AA831754
34	402.6	35.1	467	10	AV763565	AV763565
35	400.2	34.9	575	10	AV757524	AV757524
36	395.6	34.5	625	10	BB622341	BB622341
37	393.2	34.3	800	13	B1079409	B1079409
38	390.6	34.1	736	14	BM963180	BM963180
39	365.4	31.9	656	10	BE287040	BE287040
40	363	31.7	458	14	N27567	N27567
41	355.2	31.0	521	12	BF087064	BF087064
42	352.8	30.8	521	13	BE995575	BE995575
43	352.8	30.2	421	10	AV713168	AV713168
44	346.6	30.2	560	9	A1526545	A1526545
45	342.6	29.9	561	9	A1956595	A1956595

## ALIGNMENTS

RESULT 1  
LOCUS AL539324  
DEFINITION AL539324 L1 FL013 Fbri1 Homo sapiens cDNA clone CS0DF034YD22 5  
ACCESSION AL539324  
VERSION AL539324.1 GI:12868435  
KEYWORDS EST.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE  
1 (bases 1 to 934)  
Li, W.B., Gruber, C., Jesse, J. and Polayes, D.  
Full-length cDNA libraries and normalization  
Unpublished (2001)  
JOURNAL  
COMMENT  
Contact: Genoscope  
Genoscope, Centre National de Sequencage  
BP 191 91006 Evry cedex - France  
Email: segre@genoscope.cns.fr, Web : www.genoscope.cns.fr.  
Location/Qualifiers  
1. 934  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="CS0DF034YD22"  
/clone\_lib="L1 FL013 Fbri1"  
/dev\_stage="pooled tissue from post conception fetuses (20  
week 24 week and 26 week)"  
/lab\_host="DH10B"

CDNA="Organ: Fetal brain; Vector: pCMVSPORT 6; 1st strand  
/note="CDNA was primed with a NotI/collodf) primer. Five prime  
end etiched, double-stranded CDNA was digested with Not I  
and cloned into the Not I and Eco RV sites of the  
pCMVSPORT 6 vector. Library was constructed by Life  
Technologies. Contact : Feng Liang Life Technologies, a  
division of Invitrogen 9800 Medical Center Drive Rockville  
, Maryland 20850, USA Fax : (1) 301 610 8371 Email :  
fliang@lifetech.com URL :  
<http://fulllength.invitrogen.com>"

Query Match	73.9%	Score 847	DB 9	Length 934
Best Local Similarity	95.9%	Pred. No. 2e-202		
Matches	892	Conservative	4	Mismatches 31; Indels 3; Gaps 3;
QY	138	CTGCTACAAAGGTCCAGATGCGCACGTCCTCGATTTTAAACCTATGTAGATCAGGCAT	197	
Db	3	CTGGAAATATACGAGAGCGGGGAGATGATAGGGAGTTTTTAAACTATGTATGATCAGGCAT	62	
QY	198	GTAAAGCTGCTGAGAGAGTTTGTCAATATTTCTATGAGACAAATGATATAAAGAACGGG	257	
Db	63	GTAAAGCTGCTGAGAGAGTTTGTCAATATTTCTATGAGACAAATGATATAAAGAACGGG	122	
QY	258	CACATACACAGGCTGTATCTGACAAAGGCGACCTTAATTTGGAAATGGAATGCTGTTTACG	317	
Db	123	CACATACACAGGCTGTATCTGACAAAGGCGACCTTAATTTGGAAATGGAATGCTGTTTACG	182	
QY	318	GGCTGGATGCCCTAAATATTTTTTTGACACATTTGCCCTTCTAGTGTAGTCCAGGTCAATA	377	
Db	183	GGCTGGATGCCCTTAATATTTTTTTGACACATTTGCCCTTCTAGTGTAGTCCAGGTCAATA	242	
QY	378	TGTTAGATTTGCCAACCAAGTTCATGTCGACAGAACGTCATCCCAACTCAGTTCTTGATG	437	
Db	243	TGTTAGATTTGCCAACCAAGTTCATGTCGACAGAACGTCATCCCAACTCAGTTCTTGATG	302	
QY	438	TGACCAAGTGAACCTGTGAGTTTGATGAAACAAACAACATTTCTTCAACGAGAACTTCC	497	
Db	303	TGACCAAGTGAACCTGTGAGTTTGATGAAACAAACAACATTTCTTCAACGAGAACTTCC	362	
QY	498	TGCTGACTGCTGACAGTCCGACCCCAACATACGTCGTGGAAGATTCGCAAGTATGCTTCC	557	
Db	363	TGCTGACTGCTGACAGTCCGACCCCAACATACGTCGTGGAAGATTCGCAAGTATGCTTCC	422	
QY	558	GTTTTCAGAGATTTGTCTAGTAGTTTAAAGGGGCAAAAGTCCATTTCCATTTGGTCCATTAG	617	
Db	423	GTTTTCAGAGATTTGTCTAGTAGTTTAAAGGGGCAAAAGTCCATTTCCATTTGGTCCATTAG	482	
QY	618	TTCCAGCAATTTGAATTTATGTGAATTTTGTGATTTGAGAGACCTATTAATGTGCTG	677	
Db	483	TTCCAGCAATTTGAATTTATGTGAATTTTGTGATTTGAGAGACCTATTAATGTGCTG	542	
QY	678	AAACGAAATTTCTTTAATATTTTCTATTCGTCGACGACCTTTTATAGAGCTGCCAGTT	737	
Db	543	AAACGAAATTTCTTTAATATTTTCTATTCGTCGACGACCTTTTATAGAGCTGCCAGTT	602	
QY	738	TGAGACATTTGCCCTCTTAAGAGCTTTAAACAATTTTTTTTACATGCGCTATATATACATCCA	797	
Db	603	TGAGACATTTGCCCTCTTAAGAGCTTTAAACAATTTTTTTTACATGCGCTATATATACCTTCN	662	
QY	798	CTAATGACATTTCTTTAATATTTTAAACACATGATCTTTGGTACTATACATACTCACTGTG	857	
Db	663	CTAATGACATTTCTTTAATATTTTAAACACATGATCTTTGGTACTATACATACTCACTGTG	722	
QY	858	AAACGAGCTATTTGCAAAAATAAATCTTTTATATATATATATCTATGAGATGTCAGACAA	917	
Db	723	AAACGAGCTATTTGCAAAAATAAATCTTTTATATATATATATCTATGAGATGTCAGACAA	782	
QY	918	ATATTAACCTCTGGAGAGAGTGGAGTTTTTGGTTATTTAGTTTAATTTTCTAGTAATAC	977	
Db	783	ATATTAACCTCTGGAGAGAGTGGAGTTTTTGGTTATTTAGTTTAATTTTCTAGTAATAC	842	

OY	978	ACGTCGGCTGTATTTTCACTTAACAAGTGGTAATGCATTATTAATATGGCTTTTTCAAATC	1037
Dd	843	ACGTGGCTGTATTTTCAKNTAACACTGGTAATGCCATTTT-ATAATATGGCTTTTTC-AATC	900
OY	1038	AGTCACGTGAATAATAGTAGCATTTAAGTT	1067
Dd	901	AGTTCACGTG-AAATAGTAGTCAGATTTAAGTT	929

RESULT 2	978 bp	mrna	linear	EST 16-FEB-2001
AL548271				
LOCUS				
DEFINITION	AL548271			
ACCESSION	AL548271	L71.NF1006.PL2	Homo sapiens cDNA clone	CG01016TK05
VERSION	AL548271			
KEYWORDS	AL548271.1	GI:12883118		
SOURCE	EST.			
ORGANISM	human.			
REFERENCE	Homo sapiens			
AUTHORS	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
TITLE	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
JOURNAL	1. (bases 1 to 978)			
COMMENT	1. W.B., Gruber,C., Jesse,J. and Polayes,D.			
	Full-length cDNA libraries and normalization			
	Unpublished (2001)			
	Contact: Genoscope			
	Genoscope - Centre National de Sequencage			
	BP 191 91006 EVRY cedex - France			
	Email: sege@genoscope.cns.fr, web : www.genoscope.cns.fr.			
FEATURES	Location/Qualifiers			
source	1..978			

QY	41	TTTGGCGGGTTTCGCTCTCTTCATTAAGTAATGATCATTTCCGACGCGCTTGGGAGCCG-GAC	99
DB	9	TTGGAGCGGTTTCGCTCTCTTCATTAAGTAATGATCATTTCCGACGCGCTTGGGAGCCGAGAC	68
QY	100	ACGTGAGAGAGTAGTAGACGC-CGACACCTGCCAGAACACACTGCTCTCAAGTCCCGAGATGG	158
DB	69	ACGTGAGAGAGTAGTAGACGCAGACACCTGCCAGAACACACTGCTCTCAAGTCCCGAGATGG	128
QY	159	CCAGTCTCTGGATTTTAAACTATGTAGATCAGGACATGTAGAGACTGCT--GAGGACTT	216
DB	129	CCAGTCTCTGGATTTTAAACTATGTAGATCAGGACATGTAGAGACTGCTTAAAGAGATAT	188
QY	217	TGTCAATATTTACTAT-GAGACAATGATAAAGAAGACGGGACACTAACACAGCGTATC	275
DB	189	AGTCAATATTTACTATGAGACAATGATAAAGAAGACGGGACACTAACACAGCGTATC	248
QY	276	TGGACAAAGCCACCTAATATGATGTAATGCTGTTTCAGGCTGATGCCCTAATA	335
DB	249	TGGACAAAGCCACCTAATATGATGTAATGCTGTTTCAGGCTGATGCCCTAATA	308

Query Match 72.8%; Score 834; DB 9; Length 978;  
 Best Local Similarity 96.8%; Pred. No. 3,8e-199;  
 Matches 902; Conservative 2; Mismatches 22; Indels 6; Gaps 5

BASE COUNT 288 a 191 c 194 g 303 t 2 others

ORIGIN

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="CSOD1016YK06"  
 /clone\_lib="LRI.NFL006.PL2"  
 /tissue\_type="placenta"  
 /note="Vector: pCMVSPORT 6; Site\_1: NotI; 1st strand cDNA was primed with a NotI-oligo(dt) primer. Five prime end enriched, double-stranded cDNA was digested with Not I and cloned into the Not I and Eco RV sites of the pCMVSPORT 6 vector. Library was normalized. Library was constructed by Life Technologies. Contact : Feng Liang Life Technologies a division of Invitrogen, 9800 Medical Center Drive Rockville, Maryland 20850, USA Fax : (1) 301 610 8371 Email : lifang@lifetech.com URL : <http://fulllength.invitrogen.com>"

```

OY 336 ATTTTGTGACATTCCTCTAGTAGTCCAGTCAATATGTAGATTGCCAACAG 395
    |||||||
DB 309 ATTTTGTGACATTCCTCTAGTAGTCCAGTCAATATGTAGATTGCCAACAG 368
OY 396 TTGATGAGCAGCACTCAGTCCCAACTACAGTTCTTGTGTGACAGTGGAACTGTGA 455
    |||||||
DB 369 TTGATGAGCAGCACTCAGTCCCAACTACAGTTCTTGTGTGACAGTGGAACTGTGA 428
OY 456 AGTTTGTGAAACAAACAACTTCTTCAACCAAGCACTCTCTGCTGCTGAGTCCA 515
    |||||||
DB 429 AGTTTGTGAAACAAACAACTTCTTCAACCAAGCACTCTCTGCTGCTGAGTCCA 488
OY 516 CTCCCAACATACCTGTGTGGAAGATTGCAAGTGTCTCCGTTTCAAGATTGGCTGA 575
    |||||||
DB 489 CTCCCAACATACCTGTGTGGAAGATTGCAAGTGTCTCCGTTTCAAGATTGGCTGA 548
OY 576 GTAGTTAAG-GGGCAAAAGTCCATTCCTATTGGTCCATTAGTCCAGCAATTGAAATT 634
    |||||||
DB 549 GTAAAAAAGAGGCAAAAGTCCATTCCTATTGGTCCATTAGTCCAGCAATTGAAATT 608
OY 635 TATGTGAATATTTTGTATGTAGAGCACTATATATGTGTGTGAATAATTCTTTAA 694
    |||||||
DB 609 TATGTGAATATTTTGTATGTAGAGCACTATATATGTGTGTGAATAATTCTTTAA 668
OY 695 TATTTCTATTCCTGTGAGCACTTTCTAGCAGCTGCCAGTTTGGAGCAATTGCCCTTA 754
    |||||||
DB 669 TATTTCTATTCCTGTGAGCACTTTCTAGCAGCTGCCAGTTTGGAGCAATTGCCCTTA 728
OY 755 AGACCTTAAACATATTTTTCATGCTCTTATATACATTCCTAATGACATTTCTATA 814
    |||||||
DB 729 AGACCTTAAACATATTTTTCATGCTCTTATATACATTCCTAATGACATTTCTATA 788
OY 815 ATATATTAACACATATCTGTGTACTAATACATCTACCTGTAACCCAGCTTGTGAA 874
    |||||||
DB 789 ATATATTAACACATATCTGTGTACTAATACATCTACCTGTAACCCAGCTTGTGAA 848
OY 875 AATTAATATCTTTTATATATATATCTATGAGTGTGAGCAATATTAACCTGTGGAA 934
    |||||||
DB 849 AATTAATATCTTTTATATATATATCTATGAGTGTGAGCAATATTAACCTGTGGAA 908
OY 935 GAAGTGAATTTTGTGTTATAGTTAATT 966
    |||||||
DB 909 GAAGTGAATTTTGTGTTATAGTTAATT 940

RESULT 3
B0051357 993 bp mRNA linear EST 29-MAR-2002
LOCUS B0051357
DEFINITION: 5', mRNA sequence.
ACCESSION B0051357
VERSION B0051357.1 GI:19810697
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 993)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE NIH-MGC National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: DCTD/DTP/Gazdar
DNA Library Preparation: Life Technologies, Inc.
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.lnl.gov
Plate: L1AM12891 row: j column: 23
High quality sequence stop: 717.
Location/Qualifiers

```

```

source
1. .993
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5792758"
/clone_id="NIH-MGC-68"
/tissue_type="large cell carcinoma"
/lab_host="DH10B (phage-resistant)"
/notes="Organ: Lung; Vector: pCMV-Sport6; Site: 1.5 NotI;
Site 2: SalI; Cloned unidirectionally. Primer: 5' oligo dT.
Average insert size 1.8 kb. Library constructed by Life
Technologies."
BASE COUNT 306 a 174 c 192 g 321 t
ORIGIN
Query Match 71.9% Score 824.4; DB 14; Length 993;
Best Local Similarity 97.2% Pred. No. 1e-196;
Matches 861; Conservative 0; Mismatches 21; Indels 4; Gaps 2;
OY 169 GGATTTTAAACCTATATGATCAGCATGTAGAGCTGCTGAGAGTTGTCAATATTTA 228
    |||||||
DB 42 GGATTTTAAACCTATATGATCAGCATGTAGAGCTGCTGAGAGTTGTCAATATTTA 101
OY 229 CTATGAGCAATGTATTAAGAGAGGCGCACTAACCGAGCTATCTGACAAAGCCAC 288
    |||||||
DB 102 CTATGAGCAATGTATTAAGAGAGGCGCACTAACCGAGCTATCTGACAAAGCCAC 161
OY 289 CTTAATATGAAATGAAATGCTGTTTCAGGGCTGAGGCTTAATATTTTGTGACAC 348
    |||||||
DB 162 CTTAATATGAAATGAAATGCTGTTTCAGGGCTGAGGCTTAATATTTTGTGACAC 221
OY 349 ATTGCCCTCTGATGAGTTCAGGTCATATATATATGATTTGCCAAGCTGATGACAAAGC 408
    |||||||
DB 222 ATTGCCCTCTGATGAGTTCAGGTCATATATATGATTTGCCAAGCTGATGACAAAGC 281
OY 409 AACTGATCCCAAACTACAGTCTTGTGTGACCAAGTGAAGTGTGAATTTGATGAA 468
    |||||||
DB 282 AACTGATCCCAAACTACAGTCTTGTGTGACCAAGTGAAGTGTGAATTTGATGAA 341
OY 469 CAACACATTTCTTCAACCGAAGCTTCTGCTGACGTCTGATGCTCCCAACATATC 528
    |||||||
DB 342 CAACACATTTCTTCAACCGAAGCTTCTGCTGACGTCTGATGCTCCCAACATATC 401
OY 529 TGTGGAAGATGCAATGATGATGCTGCTGCTTCAAGATTTGCTAGTATTAAGGG 588
    |||||||
DB 402 TGTGGAAGATGCAATGATGATGCTGCTGCTTCAAGATTTGCTAGTATTAAGGG 461
OY 589 CAAAGTCCATTCATTTGGTCCATTAGTCCAGCAATTAATTAATTTATTTATTT 648
    |||||||
DB 462 CAAAGTCCATTCATTTGGTCCATTAGTCCAGCAATTAATTAATTTATTTATTT 521
OY 649 TGAATGTAGAACCACTATATATGTGTGGAACCTAATTTCTTAAATTTTCTATTCCT 708
    |||||||
DB 522 TGAATGTAGAACCACTATATATGTGTGGAACCTAATTTCTTAAATTTTCTATTCCT 581
OY 709 GTGAGCACTTTCTAGAGCTGCGAGCTTTGGAGATTTGCCCTTAAGCTTTAAACT 768
    |||||||
DB 582 GTGAGCACTTTCTAGAGCTGCGAGCTTTGGAGATTTGCCCTTAAGCTTTAAACT 641
OY 769 ATTTTGTACATGCTTATATATATATATATATATATATATATATATATATATAT 828
    |||||||
DB 642 ATTTTGTACATGCTTATATATATATATATATATATATATATATATATATATAT 701
OY 829 ATGATCTTGTGATACATACCTGTAAGTGAACCCAGCTTATGCAAAATTAATCTTTT 888
    |||||||
DB 702 ATGATCTTGTGATACATACCTGTAAGTGAACCCAGCTTATGCAAAATTAATCTTTT 761
OY 889 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 948
    |||||||
DB 762 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 821
OY 949 T-GGTTATAGTAAATTTTCTAGTAAACAGCTTGTGCTTTCAGTAACTGCTAA 1007
    |||||||
DB 822 TGGGTAATTAGTAAATTTTCTAGTAAACAGCTTGTGCTTTCAGTAACTGCTAA 881

```



QY 1008 TGCCATTTAAT--ATGCTTTTCAATCAGTGTGAAA 1050  
 DB 882 ATGCCATTTTAATATATAGGCTTTTCCATACAGTTTCGGGGAAA 927

RESULT 4  
 AL598693  
 LOCUS  
 DEFINITION DKFZP313A1822.1 313 (synonym: hicc2) Homo sapiens cDNA clone  
 EST.  
 ACCESSION AL598693.1 GI:15161384  
 VERSION  
 KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 793)  
 Duesterhoeft, A., Lauber, J., Mewes, H.W., Gassenhuber, J. and Wiemann  
 'S.  
 TITLE EST (Duesterhoeft, et al.)  
 JOURNAL Unpublished (1999)  
 COMMENT Contact: Duesterhoeft A  
 MIPS  
 Am Klopferseitz 18a D-82152 Martinsried, Germany  
 This is the 5' sequence of the clone insert  
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer  
 Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;  
 sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing  
 consortium of the German Genome Project.  
 No si sequence available.  
 This clone (DKFZ313A1822) is available at the RZPD in Berlin:  
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059  
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
 Location/Qualifiers  
 1..793  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="DKFZ313A1822"  
 /clone\_1lb="313 (synonym: hicc2)"  
 /dev\_stage="adult"  
 /lab\_host="DH10B"  
 /note="Vector: pTriplEx2; Site\_1: SfiIA; Site\_2: SfiIB;  
 cDNA:collection"

BASE COUNT 219 a 160 c 172 g 241 t 1 others

ORIGIN

Query Match 68.9%; Score 789.4; DB 9; Length 793;  
 Best Local Similarity 99.7%; Pred. No. 6.2e-188;  
 Matches 790; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 12 GAGATGGAGGGGTGAAAAATTTTGTGCTTTGGCGGTTTGGCTCTCTTCAATAGTAT 71  
 DB 2 GGAAGGAGGAGGTGAAAAATTTTGTGCTTTGGCGGTTTGGCTCTCTTCAATAGTAT 61

QY 72 GATCATTCGACGCCCTGGGACCGGACAGCTGAGAGTAGTGAAGCCGACAGCTGCCAG 131  
 DB 62 GATCATTCGACGCCCTGGGACCGGACAGCTGAGAGTAGTGAAGCCGACAGCTGCCAG 121

QY 132 AACACACTCTCTACAGAGTCCAGATGCGCAGCTCTCTGATTTTAAACTTTGTAGATC 191  
 DB 122 AACACACTCTCTACAGAGTCCAGATGCGCAGCTCTCTGATTTTAAACTTTGTAGATC 181

QY 192 AGGCATGTAGAGCTGCTGAGAGAGTTTGTCAATATTACTATGAGACATGATGATAAGAA 251  
 DB 182 AGGCATGTAGAGCTGCTGAGAGAGTTTGTCAATATTACTATGAGACATGATGATAAGAA 241

QY 252 GAGGGGCACTACAGAGGCTGTATCGAGAGGCAAGCCCTAATATGAGATGAGAAATGCTG 311  
 DB 242 GAGGGGCACTACAGAGGCTGTATCGAGAGGCAAGCCCTAATATGAGATGAGAAATGCTG 301

QY 312 TTTCAGGGCTGAGTCCCTAATATTTTGTGACATTCGCTTCTAGTGAATTCAGG 371

DB 302 TTTCAGGGCTGAGTCCCTAATATTTTGTGACATTCGCTTCTAGTGAATTCAGG 361

QY 372 TCAATATGTATGATGGCCACAGCTTCATGAGCAAGCACTAGTCCCAACTACAGTTC 431  
 DB 362 TCAATATGTATGATGGCCACAGCTTCATGAGCAAGCACTAGTCCCAACTACAGTTC 421

QY 432 TTGTTGTAGCAGTGGAACTGTGAAGTTGATGGAACAACAACATTTCTTCAACAG 491  
 DB 422 TTGTTGTAGCAGTGGAACTGTGAAGTTGATGGAACAACAACATTTCTTCAACAG 481

QY 492 ACTTCCTGCTAGCTCTAGTCCATCCCAACAACAGTGTGGAAGTTGCAAGTAT 551  
 DB 482 ACTTCCTGCTAGCTCTAGTCCATCCCAACAACAGTGTGGAAGTTGCAAGTAT 541

QY 552 GCTTCGCTTTCAAGATTGCTAGTATGTAAGGGGCAAAAGTCCATTCATTTGTC 611  
 DB 542 GCTTCGCTTTCAAGATTGCTAGTATGTAAGGGGCAAAAGTCCATTCATTTGTC 601

QY 612 CATTAAGTCCACAGCAATGAAATTATGCAATTTATTTGATTTGTAAGCACTAATAT 671  
 DB 602 CATTAAGTCCACAGCAATGAAATTATGCAATTTATTTGATTTGTAAGCACTAATAT 661

QY 672 GTGCTGAACATAATTTCTTATATTTTCTATTCCTGACACACTTTCTAGACGCTG 731  
 DB 662 GTGCTGAACATAATTTCTTATATTTTCTATTCCTGACACACTTTCTAGACGCTG 721

QY 732 CCAGTTTGGAGATGCCCTCCTAAGAGCTTTAAACTATTTTATACATGCTTATATAC 791  
 DB 722 CCAGTTTGGAGATGCCCTCCTAAGAGCTTTAAACTATTTTATACATGCTTATATAC 781

QY 792 ATTCACATAATG 803  
 DB 782 ATTCACATAATG 793

RESULT 5  
 AF212223  
 LOCUS Homo sapiens BM025 mRNA, complete cds.  
 DEFINITION AF212223  
 ACCESSION AF212223.1 GI:9437512  
 VERSION  
 KEYWORDS  
 SOURCE  
 ORGANISM Homo sapiens.  
 Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE 1 (bases 1 to 881)  
 Zhang, Q.H., Ye, M., Wu, X.Y., Ren, S.X., Zhao, M., Zhao, C.J., Fu, G.,  
 Shen, Y., Fan, H.Y., Lu, G., Zhong, M., Xu, X.R., Han, Z.G., Zhang, J.W.,  
 Tao, J., Huang, Q.H., Zhou, J., Hu, G.X., Gu, J., Chen, S.J. and Chen, Z.  
 Cloning and functional analysis of cDNAs with open reading frames  
 for 300 previously undefined genes expressed in CD34+ hematopoietic  
 stem/progenitor cells  
 JOURNAL Genome Res. 10 (10), 1546-1560 (2000)  
 MEDLINE 204939367  
 PUBMED 11042152  
 REFERENCE 2 (bases 1 to 881)  
 Zhao, M., Song, H., Li, N., Peng, Y., Han, Z. and Chen, Z.  
 A novel gene expressed in human bone marrow  
 JOURNAL Unpublished  
 TITLE 3 (bases 1 to 881)  
 REFERENCE  
 AUTHORS Zhao, M., Song, H., Li, N., Peng, Y., Han, Z. and Chen, Z.  
 TITLE Direct Submission  
 JOURNAL Submitted (08-DEC-1999) Chinese National Human Genome Center at  
 Shanghai, 351 Guo. Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,  
 Shanghai 201203, People's Republic of China  
 LOCATION/Qualifiers  
 1..881  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /tissue-type="bone marrow"  
 214..558

CDS



QY 791 CATTCACATATGACATCTTATATATATATTAACACATGATCTGGACTACATACAT 850  
 DB 361 CATTCACATATGACATCTTATATATATTAACACATGATCTGGACTACATACAT 420  
 QY 851 CACTGTGAACCCAGCCCTATGCAAAATTAATCTTTATATATATATCTATGAGATG 910  
 DB 421 CACTGTGAACCCAGCCCTATGCAAAATTAATCTTTATATATATATCTATGAGATG 480  
 QY 911 CAGCACAATATATACATCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 970  
 DB 481 CAGCACAATATATACATCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 540  
 QY 971 GTAAACACAGTTCCTGTTTTCAGTTTACACTGATGATGATGATGATGATGATGATG 1030  
 DB 541 GTAAACACAGTTCCTGTTTTCAGTTTACACTGATGATGATGATGATGATGATGATG 600  
 QY 1031 TCAATTCAGTTCAGTGAATATGATGATGATGATGATGATGATGATGATGATGATG 1090  
 DB 601 TCAATTCAGTTCAGTGAATATGATGATGATGATGATGATGATGATGATGATGATG 660  
 QY 1091 GAATTCAGTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 1146  
 DB 661 GAATTCAGTATGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 716  
 RESULT 7  
 LOCUS AU137118 770 bp mRNA linear EST 02-ANG-2002  
 DEFINITION AU137118 PLACE1 Homo sapiens cDNA clone IMAGE1005820 5', mRNA  
 ACCESSION AU137118  
 VERSION AU137118.1 GI:10997657  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 770)  
 Ota.T., Nishikawa.T., Suzuki.Y., Ishii.S., Saito.K., Kawai.Y.,  
 Yamamoto.J., Nakamatsu.A., Nakamura.Y., Nagai.T., Sugano.S. and  
 Isogai.T.  
 TITLE HRI human cDNA project  
 JOURNAL Unpublished (2000)  
 COMMENT Contact: Takao Isogai  
 Genomics Laboratory  
 Helix Research Institute  
 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 Tel: 81-438-52-3975  
 Fax: 81-438-52-3986  
 Email: genomics@hri.co.jp  
 HRI human cDNA project; 5'-63'-end one pass sequencing; Helix  
 Research Institute; cDNA library construction; Department of  
 Virology, Institute of Medical Science, University of Tokyo, and  
 Helix Research Institute.  
 FEATURES  
 source  
 1. 770  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_lib="PLACE1005820"  
 /clone\_id="PLACE1"  
 /tissue\_type="placenta"  
 /note="Vector: PME18SFL3"  
 BASE COUNT 237 a 157 c 139 g 232 t 5 others  
 ORIGIN  
 Query Match 61.6%; Score 706.2; DB 9; Length 770;  
 Best Local Similarity 97.1%; Pred. No. 5.3e-167;  
 Matches 748; Conservative 0; Mismatches 18; Indels 4; Gaps 3;  
 QY 130 AGAACACACTGCTACAGAGTCCC-AGATGGCCACGCTCTGGAATTTAAACTTATGATG 188  
 DB 1 AGAACACACTGCTACAGAGTCCCAGATGCGCCACGCTCTGGAATTTAAACTTATGATG 60

QY 189 ATCAGCATGTAGAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 248  
 DB 61 ATCAGCATGTAGAGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
 QY 249 GAAAGCGGGCACTAACCAAGGCTGATCTGAGCAAAAGCCCACTTATATGATGAAATG 308  
 DB 121 GAAAGCGGGCACTAACCAAGGCTGATCTGAGCAAAAGCCCACTTATATGATGAAATG 180  
 QY 309 CTGTTTCAGGGGCTGATGCTGATGATGATGATGATGATGATGATGATGATGATG 368  
 DB 181 CTGTTTCAGGGGCTGATGCTGATGATGATGATGATGATGATGATGATGATGATG 240  
 QY 369 AGGTCAATATGTTAGATGATGATGATGATGATGATGATGATGATGATGATGATG 428  
 DB 241 AGGTCAATATGTTAGATGATGATGATGATGATGATGATGATGATGATGATGATG 300  
 QY 429 TTCTTGTGTGACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 488  
 DB 301 TTCTTGTGTGACCAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 360  
 QY 489 AGAATTCCTGCTGACCTGCTGACCTGCTGACCTGCTGACCTGCTGACCTGCTGAC 548  
 DB 361 AGAATTCCTGCTGACCTGCTGACCTGCTGACCTGCTGACCTGCTGACCTGCTGAC 420  
 QY 549 ATTGCTTCGTTTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 608  
 DB 421 ATTGCTTCGTTTTCAGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
 QY 609 GTCCATTAATTCAGCAATGAAATTTATGATGATGATGATGATGATGATGATGATG 668  
 DB 481 GTCCATTAATTCAGCAATGAAATTTATGATGATGATGATGATGATGATGATGATG 540  
 QY 669 TATGCTGCAAACTAAATTTCTTATATATTTCTATCTGACAGACCTTTCTGACAG 728  
 DB 541 TATGCTGCAAACTAAATTTCTTATATATTTCTATCTGACAGACCTTTCTGACAG 600  
 QY 729 CTGCGAGTTTGAGACATGCTCCCTCAAGAGCTTTAAACTATTTTACATGCTTAA 788  
 DB 601 CTGCGAGTTTGAGACATGCTCCCTCAAGAGCTTTAAACTATTTTACATGCTTAA 660  
 QY 789 TACATTCCTCAATATGATCTTATATATATATTAACACATGATCTTGTA-CTACAT 847  
 DB 661 TACATTCCTCAATATGATCTTATATATATATTAACACATGATCTTGTA-CTACAT 720  
 QY 848 ACTCACTGTGAAGCC--AGCCTATGCAAAATTAATCTTTTATATATATATATATAT 895  
 DB 721 ACTCACTGTGAAGCCCAAGCCCAANTCCAAATTAATTAATNCNTTTTAAAAA 770  
 RESULT 8  
 LOCUS BE887056 745 bp mRNA linear EST 20-OCT-2000  
 DEFINITION BE887056  
 ACCESSION BE887056  
 VERSION BE887056  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
 1 (bases 1 to 745)  
 NIH-MGC http://mgi.nci.nih.gov/  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Unpublished (1999)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgabbs-r@mail.nih.gov  
 Tissue Procurement: ATCC  
 cDNA Library Preparation: Life Technologies, Inc.  
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1AM9719 row: h column: 11  
 High quality sequence stop: 722.

## FEATURES

source

1. 745  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:3908026"  
 /clone\_1lb="NIH\_MGC\_71"  
 /tissue\_type="leiomyosarcoma"  
 /lab\_host="DH10B (phage-resistant)"  
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1; NotI;  
 Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT.  
 Average insert size 2.1 kb."

## BASE COUNT

208 a 151 c 156 g 230 t

## ORIGIN

59.6%; Score 682.8; DB 12; Length 745;

Query Match  
 Best Local Similarity 98.6%; Pred. No. 4.1e-161;

Matches 720; Conservative 0; Mismatches 7; Indels 3; Gaps 3;

46 CGGGTTGGCTCTTCATAGTATGATCATTCGCGACCCCTGCGACCGGACAGTGA 105  
 1 CGGGTTGGCTCTTCATAGTATGATCATTCGCGACCCCTGCGACCGGACAGTGA 60  
 106 GGAGGTAGAGCGCGACACTGCCAGAACACACTGCTACAGGTCCCAATGGCCACGTC 165  
 61 GGAGGTAGAGCGCGACACTGCCAGAACACACTGCTACAGGTCCCAATGGCCACGTC 120  
 166 TCTGATTTTAAACTATATGATCAGGATGATGAGTGGTGGAGAGTTTGTCAAT 225  
 121 TCTGATTTTAAACTATATGATCAGGATGATGAGTGGTGGAGAGTTTGTCAAT 180  
 226 TTACTATGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 285  
 181 TTACTATGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240  
 286 CACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 345  
 241 CACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
 346 CACATTTGCTTCTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 405  
 301 CACATTTGCTTCTAGTATGATGATGATGATGATGATGATGATGATGATGATGAT 360  
 406 AGCAACTAGTCCCAAACTAGTCTTGTGTGACCACTGGAAGTGAAGTGTGATG 465  
 361 AGCAACTAGTCCCAAACTAGTCTTGTGTGACCACTGGAAGTGAAGTGTGATG 420  
 466 AACAACAACATTTCTTCAACCAAGATCTGCTGATGCTGATGCTGATGCTGATG 525  
 421 AACAACAACATTTCTTCAACCAAGATCTGCTGATGCTGATGCTGATGCTGATG 480  
 526 TACTGTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 585  
 481 TACTGTGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 540  
 586 GGGCAAAAGTCCATTTCTATTTGGTCCATTAATTCAGCAATGGAATTAATGATGA 645  
 541 GGGCAAAAGTCCATTTCTATTTGGTCCATTAATTCAGCAATGGAATTAATGATGA 600  
 646 TTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 705  
 601 TTTTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660  
 706 CCGTGCACAGCTTTCTAGAGCGTCCAGTGTGAGCAATGCGCTCTAGAGCTTAA 765  
 661 CCGTGCACAG-CTTTCTAGAG-7GCCAGTTTGAGCAATG-CTCTAAGAGTTAAAT 717  
 766 ACTATTTT 775  
 718 ATTTTAT 727

## RESULT 9

BI58206

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

BI58206 782 bp mRNA linear EST 21-AUG-2001  
 603199143F1 NIH\_MGC\_96 Homo sapiens cDNA clone IMAGE:5278475 5'  
 mRNA sequence.  
 BI58206  
 BI58206.1 GI:15248862  
 EST.  
 human.  
 human.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 1 (bases 1 to 782)  
 NIH-MGC <http://mgi.ncl.nih.gov/>  
 Unpublished (1999)  
 National Institutes of Health, Mammalian Gene Collection (MGC)  
 Contact: Robert Strausberg, Ph.D.  
 Email: cga@b-remail.nih.gov  
 Tissue Procurement: Miklos Palokovits, M.D., Ph.D.  
 CDNA Library Preparation: Michael J. Brownstein (NHGRI), Shiraki  
 Toshiyuki and Piero Carninci (RIKEN)  
 DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)  
 DNA Sequencing by: Incyte Genomics, Inc.  
 Clone distribution: MGC clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LLNL at:  
<http://image.llnl.gov>  
 Plate: L1AM1103 row: f column: 12  
 High quality sequence stop: 759.

## FEATURES

source

1. 782  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:5278475"  
 /clone\_1lb="NIH\_MGC\_96"  
 /tissue\_type="hypothalamus"  
 /lab\_host="DH10B"  
 /note="Organ: Brain; Vector: pBluescript (modified  
 pluescript KS+); Site: 1; BamHI; Site: 2; SalI-XhoI (gtcgag  
 ); Oligo-dT primed using primer 5'-TTTTTTTTTTTTTTVN-3',  
 size-selected for average insert size 2.3 kb and  
 normalized to 10^5. This is a primary library enriched  
 for full-length clones and constructed using the  
 cap-trapper method (Carninci, in preparation). Library  
 constructed by M. Brownstein (NIH/NHGRI, National  
 Institutes of Health). Note: this is a NIH-MGC library."

## BASE COUNT

238 a 144 c 156 g 244 t

59.5%; Score 681.4; DB 13; Length 782;

Query Match  
 Best Local Similarity 98.0%; Pred. No. 9.2e-161;

Matches 722; Conservative 0; Mismatches 11; Indels 4; Gaps 3;

169 GGATTTTAAACTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 228  
 47 GGATTTTAAACTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 106  
 229 CTATGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 288  
 107 CTATGAGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 166  
 289 CTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 348  
 167 CTTAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 226  
 349 ATTGCTTCTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 408  
 227 ATTGCTTCTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 286  
 409 AACTAGTCCCAAACTAGTCTTGTGTGACCACTGGAAGTGAAGTGTGATGATGANA 468  
 287 AACTAGTCCCAAACTAGTCTTGTGTGTGACCACTGGAAGTGAAGTGTGATGATGANA 346

	Query Match:	58.2%	Score 666.8	DB 10:	Length 721:	
	Best Local Similarity:	99.6%	Pred. No. 4.3e-157:			
	Matches 679:	Conservative	0:	Mismatches	2:	Indels 1: Gaps 1:
Oy	1	TACTTCGGGAGAGAATGSGAGGGTGGAAAATTTTGCGCTTTGCCGGCTTCCTCT	60			
Oy	9	TACTTCGGGGGGAAGATGAGGAGGTGGAAAATTTTGGCTTTGGCGGTTTTGCCTCT	68			
Oy	61	TCATAAGTATTGATTCATTCGCGACCCCTGGGMAOCCGACACGTGAGGAGGTAGTAGCACC	120			
Dd	69	TCATAAGTATTGATTCATTCGCGACCCCTGGGMAOCCGACACGTGAGGAGGTAGTAGCACC	128.			
Oy	121	GAACTGCCCAGAACACACACTCTCAAGAGTCCCAAGATGGCACGCTCTGTGATTTTAAAC	180			
Dd	129	GAACTGCCCAGAACACACACTCTCAAGAGTCCCAAGATGGCACGCTCTGTGATTTTAAAC	188			
Oy	181	TTATGTGATCAGGCATGTAGAGGCTGCTGAAGATTTGTCAATATTTACTATGACAAT	240			
Dd	189	TTATGTGATCAGGCATGTAGAGGCTGCTGAAGATTTGTCAATATTTACTATGACAAT	248			
Oy	241	GGATAAAGAAAGACGGGCACTAACGAGCTGTATCTGACAAAGCCACCTTAATATGAA	300			
Dd	249	GGATAAAGAAAGACGGGCACTAACGAGCTGTATCTGACAAAGCCACCTTAATATGAA	308			
Oy	301	TGAAAAAGCGTTTTCAGAGGCTGATGAGCCCTTAATATTTTGTGACACATTCCTTCAG	360			
Dd	309	TGAAAAAGCGTTTTCAGAGGCTGATGAGCCCTTAATATTTTGTGACACATTCCTTCAG	368			
Oy	361	TGACTTCCAGGTCAATATATGTAGATTGCCAACACAGTTTCAGACAGCACTACAGTCCA	420			
Dd	369	TGACTTCCAGGTCAATATATGTAGATTGCCAACACAGTTTCAGACAGCACTACAGTCCA	428			
Oy	421	AACATACGTTCTTGTGACACAGTGGAACTGTAAAGTTGATGAAACAAACAACATTT	480			
Dd	429	AACATACGTTCTTGTGACACAGTGGAACTGTAAAGTTGATGAAACAAACAACATTT	488			
Oy	481	CTTCACACAGAACTTCCTGCTGACTGCTCAGTCCACTCCCAACATACCTGTGTGGAAAT	540			
Dd	489	CTTCACACAGAACTTCCTGCTGACTGCTCAGTCCACTCCCAACATACCTGTGTGGAAAT	548			
Oy	541	TGCAAGATGATGCTTCCTGTTTCAAGATGTGTCTAGTAAAGGGGCAAAAGTCCATT	600			
Dd	549	TGCAAGATGATGCTTCCTGTTTCAAGATGTGTCTAGTAAAGGGGCAAAAGTCCATT	608			
Oy	601	CTCATTTGCTCCATTAGTTCCAGCAATGAATTAATGCAATTTTGTGATTGAGAG	660			
Dd	609	CTCATTTGCTCCATTAGTTCCAGCAATGA- AATTTATGTGAATTTTGTGATTGAGAG	667			
Oy	661	CACATATAATGTGCTGNAACT	682			
Dd	668	CACATATAATGTGCTGNAACT	689			
RESULT 11						
LOCUS	Bt761278	743 bp	mRNA	linear	EST 25-SEP-2001	
DEFINITION	603044031.1 NIH_MGC_116 Homo sapiens cDNA clone IMAGE:5184559 5',					
VERSION	Bt761278					
KEYWORDS	Bt761278.1 GI:15752856					
SOURCE	human.					
ORGANISM	Homo sapiens					
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;					
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.					
TITLE	NIH-MGC http://mgc.ncl.nih.gov//					
JOURNAL	National Institutes of Health, Mammalian Gene Collection (MGC)					
COMMENT	Unpublished (1999) Contact: Robert Strausberg, Ph.D. Email: rstraub@ncl.nih.gov Tissue Procurement: Life Technologies, Inc.					





```

|||||
Db 228 ATGGCTCTAGTGGATCCAGGTCATATGTTAATGTCACCAACGATTCATGACCAAGC 287
OY 409 AACCTAGCCCAACTACAGTCTTGTGTGTGACCGAGTGAAGTTGATGAGAA 468
Db 288 AACCTAGCCCAACTACAGTCTTGTGTGTGACCGAGTGAAGTTGATGAGAA 347
OY 469 CAACACATTTCTTCAACAGCAACTTCTGCTGACGCTGACCTCCCAACATATC 528
Db 348 CAACAGCATTTCTTCAACAGCAACTTCTGCTGACGCTGACCTCCCAACATATC 407
OY 529 TGTGTGGAAGATTCGACAGTATGCTGCTCCGTTTTCAGAGTGTGTAGTAAAGGG 588
Db 408 TGTGTGGAAGATTCGACAGTATGCTGCTCCGTTTTCAGAGTGTGTAGTAAAGGG 467
OY 589 CAACAGCATTTCTTCAACAGCAACTTCTGCTGACGCTGACCTCCCAACATATC 648
Db 468 CAACAGCATTTCTTCAACAGCAACTTCTGCTGACGCTGACCTCCCAACATATC 527
OY 649 TGTGTGGAAGATTCGACAGTATGCTGCTCCGTTTTCAGAGTGTGTAGTAAAGGG 708
Db 528 TGTGTGGAAGATTCGACAGTATGCTGCTCCGTTTTCAGAGTGTGTAGTAAAGGG 587
OY 709 GTCAGACCTTTTTCAGACGCTGCTGACGCTGACCTCCCAACATATC 766
Db 588 GTCAGACCTTTTTCAGACGCTGCTGACGCTGACCTCCCAACATATC 647
OY 767 TGTGTGGAAGATTCGACAGTATGCTGCTCCGTTTTCAGAGTGTGTAGTAAAGGG 824
Db 648 CAACAGCATTTCTTCAACAGCAACTTCTGCTGACGCTGACCTCCCAACATATC 707
OY 825 A-CACATGATCTTGTGA-CTACATATCTCA-CTGTGACCAAGCTATGCAAAATATA 881
Db 708 ACCACATGATCTTGTGTGACGCTGACGCTGACCTCCCAACATATC 767
OY 882 A--TCTTTTATATATTTATCTATGAGTGTGACGCAATATTAACCTCG 930
Db 768 AATCTTATATATATTTATCTATGAGTGTGACGCAATATTAACCTCG 819

RESULT 13
Bg035698 878 bp mRNA linear EST 24-JAN-2001
LOCUS 60325770F1 NIH_MGC_90 Homo sapiens cDNA clone IMAGE:4413857 5'
DEFINITION mRNA sequence.
ACCESSION Bg035698
VERSION Bg035698.1 GI:12430092
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 878)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgep@briemall.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M10140 row: 1 column: 18
High quality sequence stop: 684.
Location/Qualifiers
1. 878
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4413857"
/clone_1ib="NIH_MGC_90"

```

```

/tissue_type="adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/Note="Organ: liver; Vector: pcMV-SPORE6; Site: 1; NotI;
Site: 2; SalI; Cloned unidirectionally; oligo-dT primed.
Average insert size 1.7 kb. Library enriched for
full-length clones and constructed by Life Technologies.
Note: this is a NIH-MGC Library."
BASE COUNT 259 a 145 c 167 g 307 t
ORIGIN
Query Match 53.4%; Score 612.2; DB 12; Length 878;
Best Local Similarity 95.5%; Pred. No. 2.4e-143;
Matches 696; Conservative 0; Mismatches 23; Indels 10; Gaps 6;
OY 427 AGTCTCTGTGTGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 486
Db 1 AGTCTCTGTGTGACAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG 60
OY 487 CCAGAACTTCTGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGAC 546
Db 61 CCAGAACTTCTGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGAC 120
OY 547 TGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 606
Db 121 TGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 180
OY 607 TGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 666
Db 181 TGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 240
OY 667 AATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 726
Db 241 AATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300
OY 727 AGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 786
Db 301 AGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 360
OY 787 TATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 846
Db 361 TATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 420
OY 847 TACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 906
Db 421 TACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 480
OY 907 ATGTGACGCAATATATACACTCTGTGGGAAGAGTGTGTGTGTGTGTGTGTGTGTGT 966
Db 481 ATGTGACGCAATATATACACTCTGTGGGAAGAGTGTGTGTGTGTGTGTGTGTGTGT 539
OY 967 TCTATGTAACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1024
Db 540 TCTATGTAACACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 599
OY 1025 GCTTTTCAAAATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1078
Db 600 GCTTTTCAAAATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 659
OY 1079 TCTGACATACGTGAATGCAATATAGAG-ATGTTCAGTGTGTGTGTGTGTGTGTGT 1137
Db 660 TGGACCATACGTGAATGCAATATAGAGATGTTCAAGGCGTGTGTGTGTGTGTGTGT 719
OY 1138 TTTTGTGTGT 1146
Db 720 TTTTGTGTGT 728

RESULT 14
Bg087075 629 bp mRNA linear EST 19-OCT-2000
LOCUS Bg087075
DEFINITION PM2-H0342-130900-002-h05 H0342 Homo sapiens cDNA, mRNA sequence.
ACCESSION Bg087075
VERSION Bg087075.1 GI:10892785

```



KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
AUTHORS	Dias Neto, E., Garcia Correa, R., Veljovski-Almeida, S., Brites, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordin, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
TITLE	Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
MEDLINE	20202663
COMMENT	Contact: Simpson A.J.G. Laboratory of Cancer Genetics Ludwig Institute for Cancer Research Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil Tel: +55-11-2704922 Fax: +55-11-2707001 Email: asimpson@ludwig.org.br This sequence was derived from the FAPESP/LICR Human Cancer Genome Project. This entry can be seen in the following URL ( <a href="http://www.ludwig.org.br/scripts/gethtml2.pl?tl=612-pm2-HR0342-130">http://www.ludwig.org.br/scripts/gethtml2.pl?tl=612-pm2-HR0342-130</a> ) 900-002-H056F3-2000-09-13644-1) Seq primer: puc 18 forward High quality sequence start: 3 High quality sequence stop: 611. Location/Qualifiers 1. 629 /organism="Homo sapiens" /db_xref="taxon:9606" /clone_id="HR0342" /dev_stage="Adult" /note="Organ: head,neck; Vector: puc18; Site_1: Smal; Site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
BASE COUNT	187 a 129 c 131 g 182 t
ORIGIN	
Query Match	53.3%, Score 611, DB 12, Length 629;
Best Local Similarity	98.4%, Pred. No. 4,6e-143;
Matches 617, Conservative 0; Mismatches 10; Indels 0; Gaps 0;	
OY	119 CCGACACTGCCAGAAACACACTGCTACAAAGGCCCAAGATGGCAGCTCTCGATTTTAAA 178
Db	2 CCGACACTGCCAGAAACACACTGCTACAAAGGCCCAAGATGGCAGCTCTCGATTTTAAA 61
OY	179 ACTATATAGATTCAGGCATGTAGAGCTGCTGAGAGTTTGTCAATATTACTATAGACA 238
Db	62 ACTATATAGATTCAGGCATGTAGAGCTGCTGAGAGTTTGTCAATATTACTATAGACA 121
OY	239 ATGATTAAGAAAGACGGGCACCTAACAGGCTGTATCTGGACAAGGCCACCTTAATATGG 298
Db	122 ATGATTAAGAAAGACGGGCACCTAACAGGCTGTATCTGGACAAGGCCACCTTAATATGG 181
OY	299 AATGAATATCTGTTTACAGGCTGATGCCCTTAATATTTTGTGACATTCGCCCTTC 358
Db	182 AATGAATATCTGTTTACAGGCTGATGCCCTTAATATTTTGTGACATTCGCCCTTC 241
OY	359 AGTAGATTCAGGTCATTAATGTAGATTGGCAACAGCTTCATGACGAAGCACTAGTCC 418
Db	242 AGTAGATTCAGGTCATTAATGTAGATTGGCAACAGCTTCATGACGAAGCACTAGTCC 301
OY	419 CAAACTCACTTCTGTTGTGACACAGTGAACGTGAGTTTGATGGAACCAACACAT 478

	Db	302	CAACTACAGTTCTGTGTCACCAGTGGAACTGGGAAGTTGGATTCGAACAACAACAT	361
	Oy	479	TTCCTTCAACCGAAGACTTCTCTGTAAGTCTCAGTCCACTCCCACAACTACTGTGTGAAG	538
	Db	362	TTCTTCAACCGAAGACTTCTCTGTAAGTCTCAGTCCACTCCCACAACTACTGTGTGAAG	421
	Oy	539	ATTGCAATGATGTTGGTCCGTTTTCCAATTTGGTGCTACTACTTTAAAGGGCAAGAATGCCA	598
	Db	422	ATTGCAATGATGTTGGTCCGTTTTCCAATTTGGTGCTACTACTTTAAAGGGCAAGAATGCCA	481
	Oy	599	TTCTCATTTGGTGCATTTAGTATTCACGCAATTAATAATTTATGCAATTTATTTGATTTGAGA	658
	Db	482	TGCTCATTTGGTGCATTCAGTTCACGCAATTAATAATTTATGCAATTTATTTGATTTGAGA	541
	Oy	659	AGCACTATTAATATGTGTGTAACATAAATTTCTTAATATTTTCAATTCCTGTACGACACT	718
	Db	542	AGCACTATTAATATGTGTGTAACATAAATTTCTTAATATTTTCAATTCCTGTACGACACT	601
	Oy	719	TTTCTAGCAGCTGCCAGCTTTGGAGCAT	745
	Db	602	TTTCTAGCAGCTGCCAGCTTAGGACAT	628
RESULT 15				
BG995586	Locus	BG995586	629 bp	mRNA linear EST 13-JUN-2001
PM2-HT0342-190201-007-h05 HT0342	DEFINITION		Homo sapiens cDNA, mRNA sequence.	
BG995586	ACCESSION	BG995586		
BG995586.1 GI:14399656	VERSION			
EST.	KEYWORDS			
human.	SOURCE			
Homo sapiens	ORGANISM			
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
1 (bases 1 to 629)	REFERENCE			
Dias Neto, E., Garcia Correa, R., Verjovsky-Almeida, S., Birones, M.R.,	AUTHORS			
Nagal, M.A., da Silva, W. Jr., Zago, M.A., Bordoli, S., Costa, F.F.,				
Goldman, G.H., Carvalho, A., Matsukuma, A., Bala, G.S., Simpson, D.H.,				
Brunstein, A., deoliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare				
, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and				
Simpson, A.J.				
Shotgun sequencing of the human transcriptome with ORF expressed	TITLE			
sequence tags	JOURNAL			
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)	MEDLINE			
2020263	COMMENT			
Contact: Simpson A.J.G.				
Laboratory of Cancer Genetics				
Ludwig Institute for Cancer Research				
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,				
Brazil				
Tel: +55-11-2704922				
Fax: +55-11-2707001				
Email: asimpson@ludwig.org.br				
This sequence was derived from the FAPESP/LICR Human Cancer Genome				
project. This entry can be seen in the following URL				
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=PM2&t2=PM2-HT0342-				
190201-007-h05&t3=2001-02-19&t4=1)				
Seq primer: puc 18 forward				
High quality sequence start: 3				
High quality sequence stop: 611.				
Location/Qualifiers				
1..629				
/organism="Homo sapiens"				
/db_xref="taxon:9606"				
/clone_lib="HT0342"				
/dev_stage="Adult"				
/note="Organ: head_neck; Vector: puc18; Site_1: Smal;				
Site_2: SmaI; A mini-library was made by cloning products				
derived from ORESTES PCR (U.S. Letters Patent application				
No. 106,716 - Ludwig Institute for Cancer Research)				
profiles into the pUC 18 vector. Reverse transcription of				
tissue mRNA and cDNA amplification were performed under				
low stringency conditions."				

BASE COUNT	187 a	129 c	131 g	182 t
ORIGIN				

Query Match	53.38%	Score 611	DB 13	length 629
Best Local Similarity	98.48%	Pred. No.	4.6e-143	
Matches 617, Conservative	0	Mismatches	10	Indels 0
				Gaps 0

OY		119	CGGACACTGCCAGAACACACATGTCAACAGTCCCATGGCAGCTCTGTGATTTTAA	178
Db		2	CCGACACTGCCAAGAACACACTGTCTACAAGTCCCAGATGGCACAGCTCTGTGATTTTAA	61
OY		179	ACTTATGTATGCATCAGGCATGTAGACCTTGCTGAGAGTTTGTCAATAATTTACTATGAGACA	238
Db		62	ACTTAATGTATGCATCAGGCATGTAGACCTTGCTGAGAGTTTGTCAATAATTTACTATGAGACA	121
OY		239	ATGGATTAANAAGAAGCGGGCACTAACCAAGGCTGTATGTGGACAAGGCACCCTTAATATGG	298
Db		122	ATGGATTAANAAGAAGCGGGCACTAACCAAGGCTGTATGTGGACAAGGCACCCTTAATATGG	181
OY		299	AATGGAATGCTGTTTTCAAGGCTGAGTCCTTAATAATATTTTTTGTACACATTTGCCCTCT	358
Db		182	AATGGAATGCTGTTTTCAAGGCTGAGTCCTTAATAATATTTTTTGTACACATTTGCCCTCT	241
OY		359	AGGATATTCAGAGTCAATATGTATTAATGGCAACCAAGTCAATGAGCAAGCAACATCAGTCC	418
Db		242	AGGATATTCAGAGTCAATATGTATTAATGGCAACCAAGTCAATGAGCAAGCAACATCAGTCC	301
OY		419	CAAACTACAGTTCTTGTTGTGACACAGTGGAACTGTGAAGTTTGAATGGAACAACAACAT	478
Db		302	CAAACTACAGTTCTTGTTGTGACACAGTGGAACTGTGAAGTTTGAATGGAACAACAACAT	361
OY		479	TTCCTTAACCGAAGACTTCTCTGCTGACTGCTACATGCCAACATATCTGTGTGAG	538
Db		362	TTCCTTAACCGAAGACTTCTCTGCTGACTGCTACATGCCAACATATCTGTGTGAG	421
OY		539	ATTGCAAGATGTTGCTTCGCTTTCAAGATTGGTCTACTAATTAAAGGGCGAAAAGTCCA	598
Db		422	ATTGCAAGATGTTGCTTCGCTTTCAAGATTGGTCTACTAATTAAAGGGCGAAAAGTCCA	481
OY		599	TTCCTATTGTGTCATTAGTTCCACACAATGAAATTTATGAAATTAATTTTGATTGTAGA	658
Db		482	TGCTCATTTGTGTCATTAGTTCCACACAATGAAATTTATGAAATTAATTTTGATTGTAGA	541
OY		659	AGCACTAATAATATGTGCTGAAACTAAATTTCTTAATAATTTCTAATTCCTGTGACACCT	718
Db		542	AGCACTAATAATATGTGCTGAAACTAAATTTCTTAATAATTTCTAATTCCTGTGACACCT	601
OY		719	TTTCTTAGCAGCTCCAGTTTGGANGAT	745
Db		602	TTTCTTAGCAGCTCCAGTTTGGANGAT	628

Search completed: June 28, 2003, 03:34:25  
Job time : 1928.73 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2003, 00:27:44 ; Search time 62.6356 Seconds

(without alignments)  
5611.049 Million cell updates/sec

Title: US-09-763-902b-15

Perfect score: 1146  
1 taccctcgagagagatggtg.....tttaagtaatttttttt 1146

Scoring table: IDENTITY\_NUC

Gap 10.0 , Gapext 1.0

Searched: 443362 seqs, 153338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Issued Patents, NA:\*

1: /cgn2\_6/prodata/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/prodata/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/prodata/1/lna/5A.COMB.seq:\*  
4: /cgn2\_6/prodata/1/lna/5B.COMB.seq:\*  
5: /cgn2\_6/prodata/1/lna/PCUS.COMB.seq:\*  
6: /cgn2\_6/prodata/1/lna/Backfills1.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	40	3.5	2692	1 US-08-036-210-14	Sequence 14, Appl
C 2	40	3.5	2692	1 US-08-449-609-14	Sequence 14, Appl
C 3	39.4	3.4	1511	1 US-07-991-867B-8	Sequence 8, Appl
C 4	39.4	3.4	1511	1 US-08-107-755A-8	Sequence 8, Appl
C 5	39.4	3.4	1511	2 US-08-544-332-8	Sequence 8, Appl
C 6	39.4	3.4	1511	4 US-09-370-861A-8	Sequence 8, Appl
C 7	37	3.2	45360	4 US-09-453-702B-206	Sequence 206, App
C 8	37	3.2	45325	4 US-08-487-826B-13	Sequence 261, App
C 9	36.8	3.2	19124	4 US-09-641-638-651	Sequence 13, Appl
C 10	36.4	3.2	20674	4 US-09-122-400B-5	Sequence 5, Appl
C 11	36.2	3.2	998	4 US-08-998-416-971	Sequence 971, App
C 12	36	3.1	740	4 US-08-323-170B-1	Sequence 1, Appl
C 13	35.8	3.1	9636	4 US-08-954-441-1	Sequence 11, Appl
C 14	35.8	3.1	1251	4 US-08-991-677-11	Sequence 7, Appl
C 15	35.6	3.1	7218	1 US-08-232-463-14	Sequence 14, Appl
C 16	35.4	3.1	7218	1 US-08-232-463-14	Sequence 14, Appl
C 17	35.2	3.1	7218	1 US-08-232-463-14	Sequence 14, Appl
C 18	34.6	3.0	460	4 US-09-643-597-205	Sequence 205, Appl
C 19	34.6	3.0	1056	4 US-09-134-001C-1550	Sequence 1550, Ap
C 20	34.2	3.0	2949	4 US-09-412-554A-3	Sequence 3, Appl
C 21	34	3.0	998	4 US-09-122-400B-5	Sequence 5, Appl
C 22	34	3.0	1890	6 5312912-3	Sequence 5, Appl
C 23	33.8	2.9	7577	4 US-08-961-527-46	Sequence 46, Appl
C 24	33.8	2.9	90050	4 US-09-245-041-5	Sequence 5, Appl
C 25	33.6	2.9	2308	4 US-08-480-229C-28	Sequence 28, Appl
C 26	33.6	2.9	2308	4 US-08-659-235C-28	Sequence 28, Appl
C 27	33.6	2.9	3846	2 US-08-845-161A-5	Sequence 5, Appl

28	33.6	2.9	3846	4 US-09-270-751-5	Sequence 5, Appl
29	33.6	2.9	3846	4 US-09-168-218B-3	Sequence 3, Appl
C 30	33.6	2.9	4967	4 US-09-177-325-12	Sequence 12, Appl
C 31	33.6	2.9	4967	4 US-09-411-812A-12	Sequence 12, Appl
C 32	33.6	2.9	4967	4 US-09-590-113-12	Sequence 12, Appl
C 33	33.6	2.9	51952	3 US-08-947-823-1	Sequence 1, Appl
C 34	33.4	2.9	1004	4 US-08-714-918-69	Sequence 69, Appl
C 35	33.4	2.9	1004	4 US-09-265-315-69	Sequence 69, Appl
C 36	33.4	2.9	1004	4 US-09-265-315-69	Sequence 69, Appl
C 37	33.4	2.9	1004	4 US-09-265-417-69	Sequence 69, Appl
C 38	33.4	2.9	1029	4 US-09-134-001C-1878	Sequence 69, Appl
C 39	33.4	2.9	18443	4 US-09-078-294-6	Sequence 1878, Ap
C 40	33.4	2.9	169998	4 US-09-676-610B-24	Sequence 24, Appl
C 41	33.2	2.9	1847	4 US-09-381-849-5	Sequence 5, Appl
C 42	33.2	2.9	36159	4 US-09-749-588-3	Sequence 3, Appl
C 43	33	2.9	720	4 US-08-913-014A-4	Sequence 4, Appl
C 44	33	2.9	810	4 US-09-072-993C-5	Sequence 5, Appl
C 45	33	2.9	1353	4 US-08-913-014A-6	Sequence 6, Appl

## ALIGNMENTS

RESULT 1  
US-08-036-210-14/C  
Sequence 14, Application US/08036210  
Patent No. 5585233  
GENERAL INFORMATION:  
APPLICANT: Moller, Niels P. H.  
APPLICANT: Moller, Karin B.  
APPLICANT: Ullrich, Axel  
TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE  
PHOSPHATASE  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: PENNIE & EDMONDS  
STREET: 1155 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036-2711  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/036,210  
FILING DATE: 23-MAR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mistock, S. Leslie  
REGISTRATION NUMBER: 18,872  
REFERENCE/DOCKET NUMBER: 7683-025  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 212-790-9090  
TELEFAX: 212-869-8864/9741  
TELEX: 66141 PENNIE  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2692 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: cDNA  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 2149  
OTHER INFORMATION: /note="N-x-unknown nucleotide"  
US-08-036-210-14  
Query Match 3.5%; Score 40; DB 1; Length 2692;  
Best Local Similarity 53.9%; Pred. No. 0.09;

Matches 82: Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 987 GTTTCAGTTAAACAGTGTATGATGATTTATATATGCTTTTCAATCAGTTAGTG 1046

Db 194 GTTTCCTTTTAAAGTAAGACATTTAGTCAATATTTCTGTTAAATCAGCTCTG 135

QY 1047 AAAATAGTACAGATTAGGTTTACATACACTGACATGCAATTCGATATAGAGA 1106

Db 134 AATATGTGCAATTTATAGTACGACTCTTTCTTGTGTTAATTCATAGAACCT 75

QY 1107 TGTTCAGTGTGCTTTTTCATTTAATAATT 1138

Db 74 ATTCTACTTCTTTTCTTTTCTAATAAGTAATT 43

## RESULT 2

US-08-449-609-14/c

Sequence 14, Application US/08449609

Patent No. 5952212

GENERAL INFORMATION:

APPLICANT: Moller, Niels P.H.

APPLICANT: Moller, Karin B.

TITLE OF INVENTION: PTP-S31: A NOVEL PROTEIN TYROSINE

NUMBER OF SEQUENCES: 45

CORRESPONDENCE ADDRESS:

ADDRESS: PENNIE & EDMONDS

STREET: 1155 Avenue of the Americas

CITY: New York

STATE: New York

COUNTRY: U.S.A.

ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/449,609

FILING DATE: 24-MAY-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/036,210

FILING DATE: 23-MAR-1993

ATTORNEY/AGENT INFORMATION:

NAME: MISTOCK, S. Leslie

REGISTRATION NUMBER: 18,872

REFERENCE/DOCKET NUMBER: 7683-025

TELECOMMUNICATION INFORMATION:

TELEPHONE: 212-790-9090

TELEFAX: 212-869-8864/9741

TELEX: 66141 PENNIE

INFORMATION FOR SEQ ID NO: 14:

SEQUENCE CHARACTERISTICS:

LENGTH: 2692 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: CDNA

FEATURE:

NAME/KEY: misc\_feature

LOCATION: 2149

OTHER INFORMATION: /note="N-X-unknown nucleotide"

US-08-449-609-14

Query Match 3.5%; Score 40; DB 2; Length 2692;

Best Local Similarity 53.9%; Pred. No. 0.09;

Matches 82: Conservative 0; Mismatches 70; Indels 0; Gaps 0;

QY 987 GTTTCAGTTAAACAGTGTATGATGATTTATATATGCTTTTCAATCAGTTAGTG 1046

Db 194 GTTTCCTTTTAAAGTAAGACATTTAGTCAATATTTCTGTTAAATCAGCTCTG 135

QY 1047 AAAATAGTACAGATTAGGTTTACATACACTGACATGCAATTCGATATAGAGA 1106

Db 134 AATATGTGCAATTTATAGTACGACTCTTTCTTGTGTTAATTCATAGAACCT 75

QY 1107 TGTTCAGTGTGCTTTTTCATTTAATAATT 1138

Db 74 ATTCTACTTCTTTTCTTTTCTAATAAGTAATT 43

## RESULT 3

US-07-991-867B-8/c

Sequence 8, Application US/07991867B

Patent No. 5476781

GENERAL INFORMATION:

APPLICANT: Moyer, Richard W.

APPLICANT: Hall, Richard L.

APPLICANT: Gruidl, Michael E.

TITLE OF INVENTION: No. 5476781e1 Entomopoxvirus Expression System

NUMBER OF SEQUENCES: 66

CORRESPONDENCE ADDRESS:

ADDRESSEE: David R. Saliwanchik

STREET: 2421 N.W. 41st Street, Suite A-1

CITY: Gainesville

STATE: FL

COUNTRY: USA

ZIP: 32606

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

OPERATING SYSTEM: IBM PC compatible

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/991,867B

FILING DATE: 12-DEC-1992

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: WO 92/14818

FILING DATE: 12-FEB-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/827,685

FILING DATE: 30-JAN-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/657,584

FILING DATE: 19-FEB-1991

ATTORNEY/AGENT INFORMATION:

NAME: Saliwanchik, David R.

REGISTRATION NUMBER: 31,794

REFERENCE/DOCKET NUMBER: UF114.C3

TELECOMMUNICATION INFORMATION:

TELEPHONE: 904-375-8100

TELEFAX: 904-372-5800

INFORMATION FOR SEQ ID NO: 8:

SEQUENCE CHARACTERISTICS:

LENGTH: 1511 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: unknown

MOLECULE TYPE: DNA (genomic)

ORIGINAL SOURCE:

ORGANISM: Amsacta moorei entemopoxvirus

FEATURE:

NAME/KEY: CDS

LOCATION: complement (18..218)

FEATURE:

NAME/KEY: CDS

LOCATION: complement (234..782)

NAME/KEY: CDS

LOCATION: 852..1511

US-07-991-867B-8

Query Match 3.4%; Score 39.4; DB 1; Length 1511;



TELEPHONE: 904-375-8100  
TELEFAX: 904-372-5800  
INFORMATION FOR SEQ ID NO: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1511 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE: Amsacta moorei entomopoxvirus  
ORGANISM: Amsacta moorei entomopoxvirus  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (18..218)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (234..782)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: complement (1511..1511)  
US-08-544-332-8

Query Match 3.4%; Score 39.4; DB 2; Length 1511;  
Best Local Similarity 49.3%; Pred. No. 0.1;  
Matches 103; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 768 TATTTTTCACGCTTATATACATTCACCTAATGACATCTTATATATATTAACA 827  
DB 1013 TAAATTA 954  
QY 828 CATGATCTGTACTATACATCTGACGACCCAGCCATTCGCAAAATTAATCTTT 887  
DB 953 TATTT 894  
QY 888 TTATTTT 947  
DB 893 TAAATTT 834  
QY 948 TTGCTTATAGCTTAAATTTCTAGTAA 976  
DB 833 TTTTCATGATTAATTTTCTTGA 805

RESULT 6  
US-09-370-861A-8/c  
Sequence 8, Application US/09370861A  
Patent No. 6410221  
GENERAL INFORMATION:  
APPLICANT: MOYER, Richard W.  
APPLICANT: Hall, Richard L.  
TITLE OF INVENTION: NO. 6410221 Entomopoxvirus Expression System  
FILE REFERENCE: UFI14.C4.D1  
CURRENT APPLICATION NUMBER: US/09/370.861A  
PRIOR FILING DATE: 1999-08-09  
PRIOR APPLICATION NUMBER: US 07/991,867  
PRIOR FILING DATE: 1992-12-07  
PRIOR APPLICATION NUMBER: US 08/107,755  
PRIOR FILING DATE: 1993-08-19  
PRIOR APPLICATION NUMBER: WO 92/14818  
PRIOR FILING DATE: 1992-02-12  
PRIOR APPLICATION NUMBER: US 07/827,685  
PRIOR FILING DATE: 1992-01-30  
PRIOR APPLICATION NUMBER: US 07/657,584  
PRIOR FILING DATE: 1991-02-19  
NUMBER OF SEQ ID NOS: 78  
SOFTWARE: Patent version 3.1  
SEQ ID NO 8  
LENGTH: 1511  
TYPE: DNA  
ORGANISM: Amsacta moorei entomopoxvirus  
US-09-370-861A-8

Query Match 3.4%; Score 39.4; DB 4; Length 1511;  
Best Local Similarity 49.3%; Pred. No. 0.1;  
Matches 103; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 768 TATTTTTCACGCTTATATACATTCACCTAATGACATCTTATATATATTAACA 827  
DB 1013 TAAATTA 954  
QY 828 CATGATCTGTACTATACATCTGACGACCCAGCCATTCGCAAAATTAATCTTT 887  
DB 953 TATTT 894  
QY 888 TTATTTT 947  
DB 893 TAAATTT 834  
QY 948 TTGCTTATAGCTTAAATTTCTAGTAA 976  
DB 833 TTTTCATGATTAATTTTCTTGA 805

RESULT 7  
US-09-453-702B-206  
Sequence 206, Application US/09453702B  
Patent No. 6365723  
GENERAL INFORMATION:  
APPLICANT: Blatner, Frederick R.  
Burland, Valerie  
Perna, Nicole T.  
Plunkett, Guy  
Melch, Rod  
TITLE OF INVENTION: No. 6365723 Sequences of E. coli O157  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Plunkney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53701-2113  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 inch, 1.44MB storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 8.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/453.702B  
FILING DATE: 03-Dec-1999  
CLASSIFICATION: <unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/110,955  
FILING DATE: 04-Dec-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 251-5000  
TELEFAX: (608) 251-9166  
INFORMATION FOR SEQ ID NO: 206:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 43360  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 206:  
US-09-453-702B-206

Query Match 3.2%; Score 37; DB 4; Length 43360;  
Best Local Similarity 46.9%; Pred. No. 2.5;  
Matches 115; Conservative 0; Mismatches 130; Indels 0; Gaps 0;

OY	856	TGAACCCAGCCTATTGCAAAAATTTAAATCTTTTATATATTTATTCATGGAGATCGAGA	915
Db	22674	TGTACAGAGCTAACAAAGAAAGATTAACCTCTTTGACCTAAGATCCCTTCACAGGAGATTA	22733
OY	916	CAATATTAACACTCTGGGAAAGAGTGAAGTTTTGGTTATTAGTTAAATTTTCTAGTAA	975
Db	22734	CGAAGAAACACAGGTTGGAAGAAAGGAGAGTAAATATGTCCTCATTTATATTTCTTTCCAA	22793
OY	976	ACAGCTTGCCGTGTTTTGAGTTAAACACAGCGTAATGCCATTTAATATATGGCTTTTCAAA	1035
Db	22794	GCCTTATCTACACCCGCCGAGAAAACCGCTTAATAATAGCTGAGAAACATCCATTTTTCGGT	22853
OY	1036	TCAGTTCAAGTAAATATAGTACAGATTAGGTTTACATTAATCTGACATCTGCAAT	1095
Db	22854	GTTATTTCTTTGTTACTTTGTACACATATATCATCTTTGGGCAATCTTTAATAAGTAAACA	22913
OY	1096	GCATA	1100
Db	22914	GCATA	22918

## RESULT 8

US-09-453-702B-261  
Sequence 261, Application US/09453702B  
Patent No. 6365723  
GENERAL INFORMATION:  
APPLICANT: Blatner, Frederick R.  
Burland, Valerie  
Perna, Nicole T.  
Plunkett, Guy  
Welch, Rod  
TITLE OF INVENTION: No. 6365723el Sequences of E. coli O157  
NUMBER OF SEQUENCES: 265  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Quarles & Brady  
STREET: 1 South Pluckney Street  
CITY: Madison  
STATE: WI  
COUNTRY: US  
ZIP: 53701-2113  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.50 Inch, 1.44MB storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Word Perfect 8.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/453,702B  
FILING DATE: 03-Dec-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 60/110,955  
FILING DATE: 04-DEC-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27386  
REFERENCE/DOCKET NUMBER: 960296,95017  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 251-5000  
TELEFAX: (608) 251-9166  
INFORMATION FOR SEQ ID NO: 261:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 45325 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
SEQUENCE DESCRIPTION: SEQ ID NO: 261:  
US-09-453-702B-261

Query Match	3.28;	Score 37;	DB 4;	Length 45325;
Best Local Similarity	46.98;	Pred. NO. 2.5;		
Matches 115;	Conservative 0;	Mismatches 130;	Indels 0;	Gaps 0;

OY	856	TGAACCCAGCCTATTGCAAAAATAAAATCTTTTATAAATTTATTCATGGAGATGCAGCA	915
Db	23576	TGTACAGAGACTAACAGAAGAATTAATCCCTTTCAAGCTAATGCCCTCCAGGGAGTTA	23635
OY	916	CAATATTAACACTCTGGAGAAAGTGCGATTTTTTGGTTATTAAGTTAAATTTCTCATAAA	975
Db	23636	CGAGGAACAACGAGTTGCAAGAAATGGAGTATATATGCTCCATTTAATTTCTTCCANA	23695
OY	976	ACAGCTTGCCCTGTTTTCAAGTTAAACACGCGTAAAGCCATTTTATATATATGGCTTTTCAAA	1035
Db	23696	GCCCTTATCTACACC GCCAGAAAAACAGCTTAATAAGCTTGAGAAAATCCATTTTCCGT	23755
OY	1036	TCAAGTTAGTGAATAATAGTACAGATTTAGCTTTACATACTACTGCAATACTGCAATT	1095
Db	23756	GTTATTTCTGTGTTACTTTGTACACATATATCATTTTGGGCGAATCTTTAATAAGTGAACA	23815
OY	1096	GCATA 1100	
Db	23816	GCATA 23820	

## RESULT 9

US-08-487-8268-13/C  
: Sequence 13, Application US/08487826B  
: Patent No. 5993827  
: GENERAL INFORMATION:  
: APPLICANT: Sim, Kim L.  
: APPLICANT: Chltnis, Chetan  
: APPLICANT: Miller, Louis H.  
: APPLICANT: Peterson, David S.  
: APPLICANT: Su, Xin-zhaun  
: APPLICANT: Wellens, Thomas E.  
: TITLE OF INVENTION: BINDING DOMAINS FROM PLASMODIUM VIVAX  
: TITLE OF INVENTION: AND PLASMODIUM FALCIPARUM ERYTHROCYTE  
: NUMBER OF SEQUENCES: 45  
: CORRESPONDENCE ADDRESS:  
: ADDRESSEE: Knobbe Martens Olson & Bear  
: STREET: 620 Newport Center Drive 16th Floor  
: CITY: Newport Beach  
: STATE: California  
: COUNTRY: US  
: ZIP: 92660  
: COMPUTER READABLE FORM:  
: MEDIUM TYPE: Floppy disk  
: COMPUTER: IBM PC compatible  
: OPERATING SYSTEM: PC-DOS/MS-DOS  
: SOFTWARE: PatentIn Release #1.0, Version #1.25  
: CURRENT APPLICATION DATA:  
: APPLICATION NUMBER: US/08/487,826B  
: FILING DATE: 10-SEP-1993  
: CLASSIFICATION: 435  
: ATTORNEY/AGENT INFORMATION:  
: NAME: Israelson, Ned  
: REGISTRATION NUMBER: 29,655  
: REFERENCE/DOCKET NUMBER: NIH121.001CPI  
: TELECOMMUNICATION INFORMATION:  
: TELEPHONE: (619) 235-8550  
: TELEFAX: (619) 235-0176  
: INFORMATION FOR SEQ ID NO: 13:  
: SEQUENCE CHARACTERISTICS:  
: LENGTH: 19124 base pairs  
: type: nucleic acid  
: STRANDEDNESS: single  
: TOPOLOGY: linear  
: MOLECULE TYPE: cDNA  
: HYPOTHEetical: NO  
: ANTI-SENSE: NO  
: US-08-487-8268-13

Query Match	3.2%	Score 36.8;	DB 2;	Length 19124;
Best Local Similarity	44.5%;	Pred. No. 1.9;		
Matches 146;	Conservative	0;	Mismatches 182;	Indels 0;
			Gaps	0;



```
OY 612 CATAGTCCAGCAATGGAATTTATGTAATTTGATTTGATGAGCAAGCACTAATATAT 671
|||||
DB 6310 CATTTAAATATATATATATATATATATATATATATATATATATATATATATAT 6251
OY 672 GTGCTGAACTAAATTTCTTATATATTTTCTATTCCTGACAGCCTTTTCTAGCAGCTG 731
|||||
DB 6250 TATATTAAGAAATATTTGTATATATATATATATATATATATATATATATATAT 6191
OY 732 CCAGTTTGAGCCTTCCTCTAGAGCTTTAAACATATTTTTCATGACCTTATATAC 791
|||||
DB 6190 TATATCTTTAAAGTATATATATATATATATATATATATATATATATATATAT 6131
OY 792 ATTCACATATGATCTTATATATATATATATATATATATATATATATATATATAT 851
|||||
DB 6130 TATATATATATATATATATATATATATATATATATATATATATATATATATAT 6071
OY 852 ACTGTGACCCAGCCTATTCGAAAAATTAATCTTTTATATATATATATATATAT 911
|||||
DB 6070 ATATGATAGAAATATATATATATATATATATATATATATATATATATATAT 6011
OY 912 AGCACAATATACACCTCTGGAGAGAGT 939
|||||
DB 6010 ATATATACAAAAAGAACGACAGAGAGT 5983
|||||

RESULT 10
US-09-641-638-651
Sequence 651, Application US/09641638
Patent No. 6432648
GENERAL INFORMATION:
APPLICANT: Blumenfeld, Marta
APPLICANT: Bougueleret, Lydie
APPLICANT: Chumakov, Ilya
TITLE OF INVENTION: BI-ALLELIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING
TITLE OF INVENTION: GENES INVOLVED IN ARACHIDONIC ACID METABOLISM
FILE REFERENCE: GENSET.05ICP1
CURRENT APPLICATION NUMBER: US/09/641,638
PRIOR FILING DATE: 2000-08-16
PRIOR APPLICATION NUMBER: US 09/502,330
PRIOR FILING DATE: 2000-02-11
PRIOR APPLICATION NUMBER: US 60/133,200
PRIOR FILING DATE: 1999-05-07
PRIOR APPLICATION NUMBER: US 09/275,267
PRIOR FILING DATE: 1999-03-23
PRIOR APPLICATION NUMBER: US 60/119,917
PRIOR FILING DATE: 1999-02-12
NUMBER OF SEQ ID NOS: 1304
SOFTWARE: Patent.pm
SEQ ID NO 651.
LENGTH: 20674
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
LOCATION: 1123..3123
OTHER INFORMATION: 5' regulatory region
NAME/KEY: exon
LOCATION: 3124..3297
OTHER INFORMATION: exon 1
NAME/KEY: exon
LOCATION: 3671..4072
OTHER INFORMATION: exon 2
NAME/KEY: exon
LOCATION: 5552..5633
OTHER INFORMATION: exon 3
NAME/KEY: exon
LOCATION: 5758..5860
OTHER INFORMATION: exon 4
NAME/KEY: exon
LOCATION: 5996..6099
OTHER INFORMATION: exon 5
NAME/KEY: exon

LOCATION: 6349..6509
OTHER INFORMATION: exon 6
NAME/KEY: exon
LOCATION: 7379..7522
OTHER INFORMATION: exon 7
NAME/KEY: exon
LOCATION: 8645..8854
OTHER INFORMATION: exon 8
NAME/KEY: exon
LOCATION: 12254..12340
OTHER INFORMATION: exon 9
NAME/KEY: exon
LOCATION: 12854..13023
OTHER INFORMATION: exon 10
NAME/KEY: exon
LOCATION: 13308..13429
OTHER INFORMATION: exon 11
NAME/KEY: exon
LOCATION: 16567..16667
OTHER INFORMATION: exon 12
NAME/KEY: exon
LOCATION: 16775..16945
OTHER INFORMATION: exon 13
NAME/KEY: exon
LOCATION: 17063..17554
OTHER INFORMATION: exon 14
NAME/KEY: misc.feature
LOCATION: 17555..20674
OTHER INFORMATION: 3' regulatory region
NAME/KEY: allele
LOCATION: 1128
OTHER INFORMATION: 10-508-191 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1182
OTHER INFORMATION: 10-508-245 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1559
OTHER INFORMATION: 10-509-284 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 1570
OTHER INFORMATION: 10-509-295 : deletion of C
NAME/KEY: allele
LOCATION: 1827
OTHER INFORMATION: 10-510-173 : variable motif ATTTA or TTTT
NAME/KEY: allele
LOCATION: 2048
OTHER INFORMATION: 10-511-62 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2323
OTHER INFORMATION: 10-511-337 : insertion of T
NAME/KEY: allele
LOCATION: 2341
OTHER INFORMATION: 10-512-36 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 2623
OTHER INFORMATION: 10-512-318 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2832
OTHER INFORMATION: 10-513-250 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2844
OTHER INFORMATION: 10-513-262 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 2934
OTHER INFORMATION: 10-513-352 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 2947
OTHER INFORMATION: 10-513-365 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 3802
OTHER INFORMATION: 12-206-81 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 4062
```

Query Match	3.28;	Score 36.4;	DB 4;	Length 20674;
Best Local Similarity	46.08;	Pred. No. 2.6;		
Matches 160;	Conservative 0;	Mismatches 186;	Indels 2;	Gaps 1

Query Match	3.28;	Score 36.2;	DB 4;	Length 998;
Best Local Similarity	44.58;	Pred. No. 0.69;		
Matches 143;	Conservative	0;	Mismatches 178;	Indels 0;
			Gaps	0;

QY 678 AACCTAAATTCCTTAAATTTCTGATTCCTGACGACCTTTCTGACAGCTCCAGTT 737  
 Db 864 AAATATATTTCTTAAAGCAATTTATTTTGGTAAAACTGAAAAAACAATTAACAATAAT 805  
 QY 738 TGGAGCATTCGCCCTTAAGAGCTTAAACATATTTTTCATCGCTTATATACATTCGA 797  
 Db 804 TTCTACATAAACTGAAAAAAGAAAAATATGTTTTCAGTCTTTTACAAAAAAAACG 745  
 QY 798 CTATAGACATTCCTTAATATATATTAACAACATGATCTGTTGTACTAACATACATCAGTGG 857

Db 744 CTTAAGAACTGAAAAATATTTCTAATAATATTTTGTAAAACTGAAAAAACC 685  
Qy 858 AACCCAGCTATTCGAAAAATCTTTTATATATATCTATGAGATGTCAGACA 917  
Db 684 AGAAGATTAATTTCTAAGATTAATAAACTGAAAACTTAATATTTTAACTAAA 625  
Qy 918 ATATACACTCTGGAGAACTGAGTTTGTATATAGCTTAATTTCTAGTAAAC 977  
Db 624 AACTGAAAAAAGAAAGAAATATTTTGTAGTTTTCAAAAAATATGTTTAA 565  
Qy 978 ACGTGCCTGTTTCAGTAA 998  
Db 564 ATATTTTCAGCTTTTAA 544

## RESULT 12

US-08-998-416-971  
Sequence 971, Application US/08998416  
Patent No. 6239264

GENERAL INFORMATION:  
APPLICANT: Philippson, Peter  
APPLICANT: Pohlmann, Rainer  
APPLICANT: Steiner, Sabine  
APPLICANT: Mohr, Christine  
APPLICANT: Wendland, Jurgen  
APPLICANT: Knechtle, Philipp  
APPLICANT: Redischung, Corinne  
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSSEYII  
NUMBER OF SEQUENCES: 1152  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: No. 6239264artis Corporation  
STREET: 3054 Cornwallis Road  
CITY: Research Triangle Park  
STATE: No. 6239264th Carolina  
COUNTRY: USA  
ZIP: 27709

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/998,416  
FILING DATE: 24-DEC-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: CH 0016/97  
FILING DATE: 31-DEC-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Meigs, J. Timothy  
REGISTRATION NUMBER: 38,241  
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 919-541-8587  
TELEFAX: 919-541-8587  
INFORMATION FOR SEQ ID NO: 971:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 740 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
ORIGINAL SOURCE:  
ORGANISM: PAG1598RP  
US-08-998-416-971

Query Match 3.1%; Score 36; DB 4; Length 740;  
Best Local Similarity 47.3%; Pred. No. 0.68;  
Matches 142; Conservative 0; Mismatches 155; Indels 3; Gaps 1;

Qy 616 ACTTCACCAATGTAATTTATGTAATTTATGTAAGACCACTAATATGTC 675

Db 402 AGAACCTAATTTCAATCAATATTAATGATGATGCTGAGTCAGATCAATACATAC 461  
Qy 676 TGAAGCTAATTTCTTTAATATTTTCTATTCCTGTCAGACCTTTCTACAGCTCCAG 735  
Db 462 ATTAAGCTAATTTAATTAATTAATTTGATTTCTATATGCAATATTTATTAAT 521  
Qy 736 TTGGACATTTGCCCTTAAGAGCTTTAAACT--ATTTTTCATGCTTTATACA 792  
Db 522 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 581  
Qy 793 TTCCATTAAGCACTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 852  
Db 582 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 641  
Qy 853 CTGTGAACCGAGCTATTCGAAAAATCTTTTATATATTAATTAATTAATTAAT 912  
Db 642 TAGGAATGAGCTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 701

## RESULT 13

US-08-323-1708-1/c  
Sequence 1, Application US/083231708  
Patent No. 5733772

GENERAL INFORMATION:  
APPLICANT: Williamson, Kim C.  
APPLICANT: Kaslow, David C.  
TITLE OF INVENTION: Cloning and Expression of Plasmodium  
TITLE OF INVENTION: falciparum Transmission-Blocking Target Antigen, Pf230  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Townsend and Townsend and Crew LLP  
STREET: Two Embarcadero Center, 8th floor  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111-3834

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/323,1708  
FILING DATE: 13-OCT-1994  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/010,409  
FILING DATE: 29-JAN-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Quine, Jonathan A.  
REGISTRATION NUMBER: P-41,261  
REFERENCE/DOCKET NUMBER: 015280-113100US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 576-0200  
TELEFAX: (415) 576-0300  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9636 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 149..9556  
US-08-323-1708-1

Query Match 3.1%; Score 35.8; DB 1; Length 9636;  
Best Local Similarity 47.0%; Pred. No. 2.7;  
Matches 147; Conservative 0; Mismatches 162; Indels 4; Gaps 1;

Qy 743 CATTCCTCTTAAGAGCTTTAAACTATTTTTCATGATGCTTATATATCCACTAT 802



Mon Jun 30 08:51:38 2003

us-09-763-902b-15.rml

Page 10

Db 960 CCAATTGTATATAAACCAAAATCT 985

Search completed: June 28, 2003, 03:36:24  
Job time : 64.6356 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 27, 2003, 23:39:51 ; Search time 295.199 Seconds  
(without alignments)  
8742.532 Million cell updates/sec

Title: US-09-763-902b-15

Perfect score: 1146  
Sequence: 1 taccctccggagagaatggg.....tttaagtaattttgtttt 1146

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2185239 segs, 112599159 residues

Total number of hits satisfying chosen parameters: 4370478

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :  
1: N\_Geneseq\_101002:\*  
2: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:\*  
3: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:\*  
4: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:\*  
5: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:\*  
6: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:\*  
7: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:\*  
8: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:\*  
9: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:\*  
10: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:\*  
11: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:\*  
12: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:\*  
13: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:\*  
14: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:\*  
15: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:\*  
16: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:\*  
17: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:\*  
18: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:\*  
19: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:\*  
20: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:\*  
21: /SID52/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:\*  
22: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:\*  
23: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:\*  
24: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:\*  
25: /SID52/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1146	100.0	1146 21	AAA08041 Human protein tran
2	1136.4	99.2	2581 24	ABA05754 Human KMF2 associa
3	1082	94.4	2994 23	ABV74535 Human prostate exp
4	1064	92.8	2692 22	AAH17938 Human cDNA sequenc
5	978	85.3	1387 22	AA526214 Human genomic DNA
6	747.2	65.2	8918 22	AA526708 Human genomic DNA
7	747.2	65.2	8919 22	AA526709 Human genomic DNA
8	491.2	42.9	716 22	AAH07014 Human cDNA clone (
9	332.4	29.0	1035 23	AA575617 DNA encoding novel

10	251.6	22.0	830 21	AAA08036 Human protein tran
11	251.6	22.0	864 22	AA526215 Human cDNA encodin
12	220	19.2	245 23	ABV07047 Human prostate exp
13	220	19.2	441 23	ABV36986 Human prostate exp
14	184.8	16.1	318 23	AA558525 cDNA #1201 encodin
15	60	5.2	60 24	ABN38056 Human spliced tran
16	55.6	4.9	805 23	ABN38056 Human spliced tran
17	49.2	4.3	17738 24	ABL16401 Human immune syste
18	47.2	4.1	13123 24	ABL16401 Human immune syste
19	47.2	4.1	13123 24	ABL16401 Human immune syste
20	46.6	4.1	2932 23	ABL16400 Human immune syste
21	46.6	4.1	4050 23	ABL09616 Human immune syste
22	46.6	4.1	7883 23	ABL08308 Human immune syste
23	45.6	4.0	6062 24	ABL34079 Human immune syste
24	45.6	4.0	15161 24	ABL70458 Human immune syste
25	45.6	4.0	15161 24	AA561423 Human gene regulat
26	45.2	3.9	33053 24	ABQ67006 Human anglogenesis
27	45	3.9	6002 24	ABL32952 Human immune syste
28	45	3.9	6365 24	ABL32124 Human immune syste
29	44.8	3.9	5559 24	ABL32338 Human immune syste
30	44.4	3.9	50849 24	ABN87883 Human immune syste
31	44	3.8	6831 24	ABK28370 Human immune syste
32	44	3.8	7047 24	ABK28386 Human immune syste
33	43.8	3.8	6167 22	AA545389 DNA transcription
34	43.8	3.8	15732 22	AA545389 DNA transcription
35	43.8	3.8	15732 22	ABK28234 DNA transcription
36	43.6	3.8	9741 24	ABK33323 Human immune syste
37	43.6	3.8	47108 24	ABK31511 Signal transductio
38	43	3.8	5689 22	AA545384 Chemically pretrea
39	43	3.8	5689 22	AA545384 Chemically pretrea
40	43	3.8	5689 24	ABK28226 DNA transcription
41	42.8	3.7	11812 22	AA545501 Chemically pretrea
42	42.8	3.7	11812 22	AA545501 Chemically pretrea
43	42.8	3.7	11812 24	ABK34118 Human immune syste
44	42.8	3.7	11812 24	ABK28431 DNA transcription
45	42.8	3.7	12705 24	ABL32149 Human immune syste

## ALIGNMENTS

### RESULT 1

ID AAA08041 standard; cDNA: 1146 BP.

AC AAA08041;

XX 19-JUN-2000 (first entry)

DE Human protein transport molecule (PRM) encoding cDNA seq ID NO:15.

XX Human; protein transport molecule; PRM; diagnosis; cytostatic;

XX antihypertensive; antidiabetic; antipneumatic; antipneumatic; osteoporotic;

XX antiallergic; antidiabetic; antipneumatic; antipneumatic; osteoporotic;

XX antineoplastic; antineoplastic; antipneumatic; antipneumatic; osteoporotic;

XX antineoplastic; antineoplastic; antipneumatic; antipneumatic; osteoporotic;

XX antineoplastic; antineoplastic; antipneumatic; antipneumatic; osteoporotic;

XX antineoplastic; antineoplastic; antipneumatic; antipneumatic; osteoporotic;

XX antineoplastic; antineoplastic; antipneumatic; antipneumatic; osteoporotic;

XX antineoplastic; antineoplastic; antipneumatic; antipneumatic; osteoporotic;

XX antineoplastic; antineoplastic; antipneumatic; antipneumatic; osteoporotic;

XX antineoplastic; antineoplastic; antipneumatic; antipneumatic; osteoporotic;

XX antineoplastic; antineoplastic; antipneumatic; antipneumatic; osteoporotic;

XX antineoplastic; antineoplastic; antipneumatic; antipneumatic; osteoporotic;

New human protein transport-associated polypeptide and polynucleotide useful for diagnosis, prevention and treatment of cell proliferative and secretory disorders such as leukemia, cystic fibrosis

Claim 9; Page 74; 75pp: English.

AAA08035 to AAA08042 encode the human protein transport-associated molecules (PRAMs) given in AYR82317 to AYR82324. The PRAMs have cytoskeletal, antirheumatic, antifibrotic, immunosuppressant, antiarteriosclerotic, antiallergic, antidiabetic, antileukemic, antineuritic, osteopathic, dermatological, antianemic, antipsoriatic, hepatotropic, antigout, antiinflammatory and antiviral activities, and regulate protein transport. PRAM proteins and antagonists are useful for preventing or treating a disorder associated with decreased or increased expression or activity of PRAM. PRAM polynucleotides are useful for diagnosing conditions associated with PRAM, comprising detecting PRAM by forming a hybridisation complex, preferably after PCR amplifying the biological sample. Diseases prevented, treated or diagnosed include cell proliferative disorders such as cancers, immune disorders, secretory disorders and other conditions associated with abnormal vesicle trafficking, such as allergies, asthma, urticaria and autoimmune haemolytic anaemia. Anti-PRAM antibodies may be used as antagonists, as a targeting or delivery mechanism for bringing pharmaceutical agents into cells or tissues expressing PRAM and for diagnosis of PRAM-related disorders. PRAM, its catalytic or immunogenic fragments are useful for drug screening using libraries of compounds. PRAM polynucleotides are useful for generating hybridisation probes useful in mapping the naturally occurring genomic sequences.

SQ Sequence 1146 BP; 331 A; 212 C; 228 G; 375 T; 0 other;

Query Match 100.0%; Score 1146; DB 21; Length 1146;  
Best Local Similarity 100.0%; Pred. No. 4, 2e-291;  
Matches 1146; Conservative 0; Mismatches 0; Indels 0; Gaps 0

OY 1 TACTTCGGGAGAGAATGGGAGGTGGAATAATTCTTGCGTTTGGGCGGTTCCGCTCTC 60  
DB 1 TAC TTC CGG GA AGA AT GG GAG GTT GGA ATA ATT CT TG CG TT T GG G CG GT TC CG CT CT C 60  
OY 61 TCATTAAGTATGATCATTCGCCGAGCCCTGGGACCAGGCAGTGTGAGGAGTGTGACGCC 120  
DB 61 TCATTAAGTATGATCATTCGCCGAGCCCTGGGACCAGGCAGTGTGAGGAGTGTGACGCC 120  
OY 121 GACACTGCCAGAACACACTGCTACAAGGTCGCCAGATGGCACGTCCTGTGATTTAAAC 180  
DB 121 GACACTGCCAGAACACACTGCTACAAGGTCGCCAGATGGCACGTCCTGTGATTTAAAC 180  
OY 181 TTATGTAGTCAAGGCATGTAGAGCTGCTGAGAGATTGTCAATATTACTATGACCAAT 240  
DB 181 TTATGTAGTCAAGGCATGTAGAGCTGCTGAGAGATTGTCAATATTACTATGACCAAT 240  
OY 241 GGATAAAAAARACGGGCACTAACAGGCTGTATCGGCAAAGGCCACCTTATATGAA 300  
DB 241 GGATAAAAAARACGGGCACTAACAGGCTGTATCGGCAAAGGCCACCTTATATGAA 300  
OY 301 TGGAAATGCTGTTTCAAGGCGCTGGATGGCCTTAATATATTTTTTGACACATTTGCTCTAG 360  
DB 301 TGGAAATGCTGTTTCAAGGCGCTGGATGGCCTTAATATATTTTTTTTGACACATTTGCTCTTAG 360  
OY 361 TAGATTCCAAGTCAATATGTATGATTCACACAGTTCATGAGCAAGCAACTAGTCCC 420  
DB 361 TAGATTCCAAGTCAATATGTATGATTCACACAGTTCATGAGCAAGCAACTAGTCCC 420  
OY 421 AACCTACAGTTCTGTTGTGACCAAGTGAACCTGTGAAGTTGATGGAACAAACACATTT 480  
DB 421 AACCTACAGTTCTGTTGTGACCAAGTGAACCTGTGAAGTTGATGGAACAAACACATTT 480  
OY 481 CTTCACACGAACCTTCCTGCTACGCTGCTAGTCACCTCCCAACAAATACGTGTGGAGAT 540

[illegible]





(a) assessing whether a patient is afflicted with prostate cancer;  
 (b) monitoring the progression of prostate cancer in a patient;  
 (c) assessing the efficacy of a test compound to inhibit prostate cancer in a patient;  
 (d) assessing the efficacy of a therapy for inhibiting prostate cancer in a patient;  
 (e) selecting a composition for inhibiting prostate cancer in a patient;  
 (f) assessing the prostate cell carcinogenic potential of a compound;  
 (g) determining whether prostate cancer has metastasized in a patient;  
 (h) assessing the aggressiveness or indolence of prostate cancer in a patient;  
 (i) is also useful as a pharmacodynamic or pharmacogenomic marker.

Sequence 2994 BP: 977 A; 535 C; 739 G; 730 T; 13 other:

Query Match 94.4%; Score 1082; DB 23; Length 2994;  
 Best Local Similarity 99.9%; Pred. No. 4e-274;  
 Matches 1093; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

53 CGCTCTCTCATAGATGATGATCATTCGCGACCCGCGGACGAGACGTGAGAGTA 112  
 |||||||  
 8 CGCTCTCTCATAGATGATGATCATTCGCGACCCGCGGACGAGACGTGAGAGTA 67  
 |||||||  
 113 GTGACGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 172  
 |||||||  
 68 GTGACGCGGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 126  
 |||||||  
 173 TTTAAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 232  
 |||||||  
 127 TTTAAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 186  
 |||||||  
 233 GAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 292  
 |||||||  
 187 GAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 246  
 |||||||  
 293 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 352  
 |||||||  
 247 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 306  
 |||||||  
 353 CCTTCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 412  
 |||||||  
 307 CCTTCTAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 366  
 |||||||  
 413 CAGTCCCAACTACATCTTGTGTGACAGTGAAGTGAAGTGAAGTGAAGTGAAGT 472  
 |||||||  
 367 CAGTCCCAACTACATCTTGTGTGACAGTGAAGTGAAGTGAAGTGAAGTGAAGT 426  
 |||||||  
 473 CAACATTTCTTCAACAGAACTTCTGTGACTGTGACTGTGACTGTGACTGTGACT 532  
 |||||||  
 427 CAACATTTCTTCAACAGAACTTCTGTGACTGTGACTGTGACTGTGACTGTGACT 486  
 |||||||  
 533 TGGAAATGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 592  
 |||||||  
 487 TGGAAATGGAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 546  
 |||||||  
 593 AGTCATTTCTCATTTGTCATTTGTCATTTGTCATTTGTCATTTGTCATTTGTCAT 652  
 |||||||  
 547 AGTCATTTCTCATTTGTCATTTGTCATTTGTCATTTGTCATTTGTCATTTGTCAT 606  
 |||||||  
 653 TGTGAAGACATATATATGATGATGATGATGATGATGATGATGATGATGATGAT 712  
 |||||||  
 607 TGTGAAGACATATATATGATGATGATGATGATGATGATGATGATGATGATGAT 666  
 |||||||  
 713 GCACCTTTCTGACAGCTGACAGCTGACAGCTGACAGCTGACAGCTGACAGCTTAA 772  
 |||||||  
 667 GCACCTTTCTGACAGCTGACAGCTGACAGCTGACAGCTGACAGCTGACAGCTTAA 726  
 |||||||  
 773 TTTTACATGCTAT 832  
 |||||||  
 727 TTTTACATGCTAT 786  
 |||||||  
 833 TCTTGTACTAATACATGCTGCTGACAGCTGCTGACAGCTGCTGACAGCTTATTA 892  
 |||||||

DB 787 TCTTGTACTAATACATGCTGCTGACAGCTGCTGACAGCTGCTGACAGCTTATTA 846  
 |||||||  
 OY 893 AATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 952  
 |||||||  
 DB 847 AATATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 906  
 |||||||  
 OY 953 TATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1012  
 |||||||  
 DB 907 TATTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 966  
 |||||||  
 OY 1013 TTTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1072  
 |||||||  
 DB 967 TTTTAAATATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1026  
 |||||||  
 OY 1073 AACTACTGACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1132  
 |||||||  
 DB 1027 AACTACTGACATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1086  
 |||||||  
 OY 1133 GTAAATTTTGT 1146  
 |||||||  
 DB 1087 GTAAATTTTGT 1100  
 |||||||

RESULT 4  
 AAH17938  
 ID AAH17938 standard; CDNA: 2692 BP.

AAH17938;

26-JUN-2001 (first entry)

Human cDNA sequence SEQ ID NO:17706.

Human; primer: detection; diagnosis; antisense therapy; gene therapy; ss.

Homo sapiens.

EP1074617-A2.

07-FEB-2001.

28-JUL-2000; 2000EP-0116126.

29-JUL-1999; 99JP-0248036.

27-AUG-1999; 99JP-0300253.

11-JAN-2000; 2000JP-0118776.

02-MAY-2000; 2000JP-0183767.

09-JUN-2000; 2000JP-0241899.

(HELI-) HELIX RES INST.

Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;

Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;

WPI: 2001-318749/34.

Claim 8; SEQ ID 17706; 2537pp + CD ROM; English.

The present invention describes primer sets for synthesizing 5602 full-length cDNAs defined in the specification. Where a primer set comprises: (a) an oligo-dT primer and an oligonucleotide complementary to the complementary strand of a polynucleotide which comprises one of the 5602 nucleotide sequences defined in the specification, where the oligonucleotide comprises at least 15 nucleotides; or (b) a combination of an oligonucleotide comprising a sequence complementary to the complementary strand of a polynucleotide which comprises a 5'-end sequence and an oligonucleotide comprising a sequence complementary to a polynucleotide which comprises a 3'-end sequence, where the

oligonucleotide comprises at least 15 nucleotides and the combination of the 5'-end sequence/3'-end sequence is selected from those defined in the specification. The primer sets can be used in antisense therapy and in gene therapy. The primers are useful for synthesizing polynucleotides, particularly full-length cDNAs. The primers are also useful for the detection and/or diagnosis of the abnormality of the proteins encoded by the full-length cDNAs. The primers allow obtaining of the full-length cDNAs easily without any specialised methods. AAH03166 to AAH1628 and AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to AAH95893 represent human amino acid sequences; and AAH13629 to AAH16332 represent oligonucleotides, all of which are used in the exemplification of the present invention.

Sequence 2692 BP; 853 A; 471 C; 478 G; 890 T; 0 other.

Query Match 92.8%; Score 1064; DB 22; Length 2692;  
Best Local Similarity 100.0%; Pred. No. 2e-269;  
Matches 1064; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 83 AGCCCTGCGGACGCGACGAGAGTAGTGACGCCGACACGCCAGAACACACTGCT 142
DB 196 AGCCCTGCGGACGCGACGAGAGTAGTGACGCCGACACGCCAGAACACACTGCT 255
QY 143 ACAAGCTCCAGATGCCAGCTCTGTGATTTTAAACTTATGTAGATCAGCATGTAGA 202
DB 256 ACAAGTCCAGATGCCAGCTCTGTGATTTTAAACTTATGTAGATCAGCATGTAGA 315
QY 203 GCTGCTGAGAGAGTTTGTCAATATTACTATGAGACAATGTATAAGAGAGGACACTA 262
DB 316 GCTGCTGAGAGAGTTTGTCAATATTACTATGAGACAATGTATAAGAGAGGACACTA 375
QY 263 ACCAGGCTGATCTGGACAAGGCCACTTAATATGAAATGGAATGCTGTTACAGGGCTG 322
DB 376 ACCAGGCTGATCTGGACAAGGCCACTTAATATGAAATGGAATGCTGTTACAGGGCTG 435
QY 323 GATGCCCTAATAATTTTTTTGACACATGCTCTGTAGTGTCCAGGTCATATGTTA 382
DB 436 GATGCCCTAATAATTTTTTTGACACATGCTCTGTAGTGTCCAGGTCATATGTTA 495
QY 383 GATTGCCACACAGTTCATGAGCAAGCAACATGCTCCAAATACAGTTCTTGTGTGACC 442
DB 496 GATTGCCACACAGTTCATGAGCAAGCAACATGCTCCAAATACAGTTCTTGTGTGACC 555
QY 443 AGTGGAACTGTGAAGTTTGAATGAAACAACAATTTCTTCAACCCAGAACTTCTGTG 502
DB 556 AGTGGAACTGTGAAGTTTGAATGAAACAACAATTTCTTCAACCCAGAACTTCTGTG 615
QY 503 ACTGCTCAGTCACTCCCAACAATAGTGTGGAAGATTGCAATGCTTCCGTTT 562
DB 616 ACTGCTCAGTCACTCCCAACAATAGTGTGGAAGATTGCAATGCTTCCGTTT 675
QY 563 CAAGATTGGTCTAGTACTTAAAGGGGCAAAAGTCCATCTCATTTGGTCAATTAGTTCA 622
DB 676 CAAGATTGGTCTAGTACTTAAAGGGGCAAAAGTCCATCTCATTTGGTCAATTAGTTCA 735
QY 623 GCATTTAAATTTATGTGAATTTATTTGATTGTGAAGCACTATAATATGTGCTGAACT 682
DB 736 GCATTTAAATTTATGTGAATTTATTTGATTGTGAAGCACTATAATATGTGCTGAACT 795
QY 683 AAATTTCTTAATATTTTCTATCTCTGACGACCTTTTCTAGAGCTGCCAGTTTGGAG 742
DB 796 AAATTTCTTAATATTTTCTATCTCTGACGACCTTTTCTAGAGCTGCCAGTTTGGAG 855
QY 743 CATTGCCCTTAAGAGCTTAAACATTTTCTTACATGCTTATATACATTCACCTAAT 802
DB 856 CATTGCCCTTAAGAGCTTAAACATTTTCTTACATGCTTATATACATTCACCTAAT 915
QY 803 GACATCTTATTAATATTAATTAACACATGATCTTGTACTAATACACTGCTGGAACCC 862
DB 916 GACATCTTATTAATATTAATTAACACATGATCTTGTACTAATACACTGCTGGAACCC 975
QY 863 AGCCTATGCAAAATTAATCTTTTATATATATATATATATATATATATATATATATAT 922

```

```

DB 976 AGCCTATGCAAAATTAATATCTTTTATATATATATATATATATATATATATATATAT 1035
QY 923 AACTCTGGGAAGAAGTGGAGTCTTTTGGTATATAGTTAATTTTCTAGTAAACACGTT 982
DB 1036 AACTCTGGGAAGAAGTGGAGTCTTTTGGTATATAGTTAATTTTCTAGTAAACACGTT 1095
QY 983 GCCGTGTTTCAGTAAACCTGCTAATGCCATTTTAATATATAGCTTTTCAAAACAGTTC 1042
DB 1096 GCCGTGTTTCAGTAAACCTGCTAATGCCATTTTAATATATAGCTTTTCAAAACAGTTC 1155
QY 1043 AGTGAATAATAGTACAGATTAGTTAGTTTACATACACTGACATACATGCAATTCATATA 1102
DB 1156 AGTGAATAATAGTACAGATTAGTTAGTTTACATACACTGACATACATGCAATTCATATA 1215
QY 1103 GAGATGTCAGTGTGCTTTTTCATTTTAAGTAAATTTTGTGTTT 1146
DB 1216 GAGATGTCAGTGTGCTTTTTCATTTTAAGTAAATTTTGTGTTT 1259

RESULT 5
AAS26214
ID AAS26214 standard; cDNA; 1387 BP.
XX
AC AAS26214;
XX
DT 07-NOV-2001 (first entry)
XX
DE Human cDNA encoding a novel secreted protein, Seq ID 393.
XX
KW Human; immunosuppressive; antiarthritic; ss. antirheumatic;
KW cytosolic; cardiant; vasotropic; cerebroprotective; nootropic;
KW neuroprotective; antibacterial; virocidic; fungicide; ophthalmological;
KW vulnary; secreted protein; rheumatoid arthritis;
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;
KW cerebrovascular disorder; cerebral ischemia; angiogenesis;
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;
KW corneal infection; wound healing; epithelial cell proliferation;
KW skin ageing; food additive; preservative; antiproliferative.
XX
OS Homo sapiens.
XX
PN M0200155322-A2.
XX
PD 02-AUG-2001.
XX
PF 17-JAN-2001; 2001M0-US01341.
XX
PR 31-JAN-2000; 2000US-0179065.
PR 04-FEB-2000; 2000US-0180628.
PR 24-FEB-2000; 2000US-0184664.
PR 02-MAR-2000; 2000US-0186350.
PR 16-MAR-2000; 2000US-0189874.
PR 17-MAR-2000; 2000US-0190076.
PR 18-APR-2000; 2000US-0198123.
PR 19-MAY-2000; 2000US-0205515.
PR 07-JUN-2000; 2000US-0209467.
PR 28-JUN-2000; 2000US-0214886.
PR 30-JUN-2000; 2000US-0215135.
PR 07-JUL-2000; 2000US-0216647.
PR 07-JUL-2000; 2000US-0216880.
PR 11-JUL-2000; 2000US-0217487.
PR 11-JUL-2000; 2000US-0217496.
PR 14-JUL-2000; 2000US-0218290.
PR 26-JUL-2000; 2000US-0220963.
PR 26-JUL-2000; 2000US-0220964.
PR 14-AUG-2000; 2000US-0224518.
PR 14-AUG-2000; 2000US-0224519.
PR 14-AUG-2000; 2000US-0225213.
PR 14-AUG-2000; 2000US-0225214.
PR 14-AUG-2000; 2000US-0225266.
PR 14-AUG-2000; 2000US-0225267.
PR 14-AUG-2000; 2000US-0225268.
PR 14-AUG-2000; 2000US-0225270.

```







Db 2079 ACTCCACACATATCTGTGGAGATTCAGATGATTCCTTCACGATTGGCT 2020  
QY 575 AGAGTAAAGGGGCAAAAGCCATTCATTTGGTCCATTAGTCCAGCAATGAATT 634  
Db 2019 AGAGTTAAAGGGGCAAAAGCCATTCATTTGGTCCATTAGTCCAGCAATGAATT 1960  
QY 635 TATGCAATATATTTGATTTAGAGCAGATATATATGCTGGAACATAATTTCTTAA 694  
Db 1959 TATGCAATATATTTGATTTAGAGCAGATATATATGCTGGAACATAATTTCTTAA 1900  
QY 695 TATTTCTATTCCTGTCAGACCTTTTCTAGACAGCTCCAGATTGGAGCATTTGCCCTTA 754  
Db 1899 TATTTCTATTCCTGTCAGACCTTTTCTAGACAGCTCCAGATTGGAGCATTTGCCCTTA 1840  
QY 755 AGAGCTTAAACATATTTTTCATATGCTTATATATTCATATTCATATGACATCTTATA 814  
Db 1839 AGAGCTTAAACATATTTTTCATATGCTTATATATTCATATTCATATGACATCTTATA 1780  
QY 815 ATATATTAACACATATCTGTGTAACATACATCTGTAACCCAGCCATTTGCA 874  
Db 1779 ATATATTAACACATATCTGTGTAACATACATCTGTAACCCAGCCATTTGCA 1720  
QY 875 AAATAAATCTTTTATATATATTCATGATGAGCAGCAATATACACTGCGAA 934  
Db 1719 AAATAAATCTTTTATATATATTCATGATGAGCAGCAATATACACTGCGAA 1660  
QY 935 GAAGTGAGTTTGTGTTATAGGTAATTTCTAGTAAACAGCTGCTGTTTTCAG 994  
Db 1659 GAAGTGAGTTTGTGTTATAGGTAATTTCTAGTAAACAGCTGCTGTTTTCAG 1600  
QY 995 TTAACACTGTAAGCCATTTTATATATGCTTTTCAATGCTGAGGAAATAGT 1054  
Db 1599 TTAACACTGTAAGCCATTTTATATATGCTTTTCAATGCTGAGGAAATAGT 1540  
QY 1055 ACAGATTAGTTTACATCTGACATCTGACATCTGGAATGATAGATTTGAGT 1114  
Db 1539 ACAGATTAGTTTACATCTGACATCTGACATCTGGAATGATAGATTTGAGT 1480  
QY 1115 GGTGCTTTTCATTTAAGTAATTTTGTGTTT 1146  
Db 1479 GGTGCTTTTCATTTAAGTAATTTTGTGTTT 1448

RESULT 7  
AAS26709/c  
ID AAS26709 standard; DNA; 8919 BP.  
XX AAS26709;  
AC  
XX  
XX 07-NOV-2001. (first entry)  
DT  
DE Human genomic DNA encoding partial novel secreted protein, Seq ID 1683.  
XX  
XX  
XX Human: immunosuppressive; antiarthritic; ds; antihemetic;  
KM cytoskeletal; cardiant; vasotropic; cerebroprotective; neurologic;  
KM neuroprotective; antibacterial; virucide; fungicide; ophthalmologic;  
KM vulnery; secreted protein; rheumatoid arthritis;  
KM hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KM cerebrovascular disorder; cerebral ischemia; angiogenesis;  
KM nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KM corneal infection; wound healing; epithelial cell proliferation;  
KM skin aging; food additive; preservative; antiproliferative.  
XX  
XX Homo sapiens.  
OS  
XX  
XX W0200155322-A2.  
PN  
XX  
XX 02-AUG-2001.  
PD  
XX  
XX 17-JAN-2001; 2001MO-US01341.  
PF  
XX  
XX 31-JAN-2000; 2000US-0179065.  
PR

PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184654.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.  
PR 18-AUG-2000; 2000US-0226279.  
PR 22-AUG-2000; 2000US-0226681.  
PR 22-AUG-2000; 2000US-0226868.  
PR 22-AUG-2000; 2000US-0227182.  
PR 23-AUG-2000; 2000US-0227009.  
PR 30-AUG-2000; 2000US-0228924.  
PR 01-SEP-2000; 2000US-0229287.  
PR 01-SEP-2000; 2000US-0229343.  
PR 01-SEP-2000; 2000US-0229344.  
PR 01-SEP-2000; 2000US-0229345.  
PR 05-SEP-2000; 2000US-0229509.  
PR 05-SEP-2000; 2000US-0229513.  
PR 06-SEP-2000; 2000US-0230437.  
PR 06-SEP-2000; 2000US-0230438.  
PR 08-SEP-2000; 2000US-0231242.  
PR 08-SEP-2000; 2000US-0231243.  
PR 08-SEP-2000; 2000US-0231244.  
PR 08-SEP-2000; 2000US-0231413.  
PR 08-SEP-2000; 2000US-0231414.  
PR 08-SEP-2000; 2000US-0232080.  
PR 08-SEP-2000; 2000US-0232081.  
PR 12-SEP-2000; 2000US-0231968.  
PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 14-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0234984.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236327.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0236802.  
PR 02-OCT-2000; 2000US-0237037.







DR WPI: 2001-639362/73.  
 DR P-PSDB: ABG11430.  
 PT New isolated polynucleotide and encoded polypeptides, useful in  
 PT diagnostics, forensics, gene mapping, identification of mutations  
 PT responsible for genetic disorders or other traits and to assess  
 PT biodiversity  
 PS Claim 1; SEQ ID No 11421; 103pp; English.

CC The invention relates to isolated polynucleotide (I) and  
 CC polypeptide (II) sequences. (I) is useful as hybridisation probes,  
 CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome  
 CC and gene mapping, and in recombinant production of (II). The  
 CC polynucleotides are also used in diagnostics as expressed sequence tags  
 CC for identifying expressed genes. (I) is useful in gene therapy techniques  
 CC to restore normal activity of (II) or to treat disease states involving  
 CC (II). (II) is useful for generating antibodies against it, detecting or  
 CC quantitating a polypeptide in tissue, as molecular weight markers and as  
 CC a food supplement. (II) and its binding partners are useful in medical  
 CC imaging of sites expressing (II). (I) and (II) are useful for treating  
 CC disorders involving aberrant protein expression or biological activity.  
 CC The polypeptide and polynucleotide sequences have applications in  
 CC diagnostics, forensics, gene mapping, identification of mutations  
 CC responsible for genetic disorders or other traits to assess biodiversity  
 CC and to produce other types of data and products dependent on DNA and  
 CC amino acid sequences. AAS64197-AAS94564 represent novel human  
 CC diagnostic coding sequences of the invention.  
 CC Note: The sequence data for this patent did not appear in the printed  
 CC specification, but was obtained in electronic format directly from WIPRO  
 CC at ftp.wipro.int/pub/published\_pcl\_sequences.

CC Sequence 1035 BP; 318 A; 219 C; 222 G; 276 T; 0 other:

Query Match 29.0%; Score 332.4; DB 23; Length 1035;

Best Local Similarity 98.9%; Pred. No. 2.6e-77;  
 Matches 356; Conservative 0; Mismatches 1; Indels 3; Gaps 2;

QY 15 AATGGGAGGTGGGAAATTTTGTGCGTTGGCGGTTCCCTCTCTATAGATTTCAT 74  
 DB 55 ACTGGGAGGTGGGAAATTTTGTGCGTTGGCGGTTCCCTCTCTATAGATTTCAT 114  
 QY 75 CATTCCGAGGCGCTGCGGACCGGACACCTGAGAGAGTGTACCGCCGACCTGCCAGAAC 134  
 DB 115 CATTCCGAGGCGCTGCGGACCGGACACCTGAGAGAGTGTACCGCCGACCTGCCAGAAC 174  
 QY 135 ACACCTGTACAGAGTCCAGATGGCCAGCTCTGTGATTATAAACTTATGTAGATCAGG 194  
 DB 175 ACACCTGTACAGAGTCCAGATGGCCAGCTCTGTGATTATAAACTTATGTAGATCAGG 234  
 QY 195 CATGTAGAGCTGTGAGAGAGTTTGCATATTCTATAGAGACAATGATTAAGAAACAC 254  
 DB 235 CATGTAGAGCTGTGAGAGAGTTTGCATATTCTATAGAGACAATGATTAAGAAACAC 294  
 QY 255 GGGGCACTAACAGAGGTGTATGTGACAGGCGCACCCTTATATGATGAAATGCTGTTT 314  
 DB 295 GGGGCACTAACAGAGGTGTATGTGACAGGCGCACCCTTATATGATGAAATGCTGTTT 354  
 QY 315 CAGGCGCTGATGCCCTTAATTA--TTTTTTGACACATGTCCTTC--TAGTGAAGTCCAGG 371  
 DB 355 CAGGCGCTGATGCCCTTAATTAATTTTTTTTGGACACATGTCCTTC--TAGTGAAGTCCAGG 414

RESULT 10  
 ID AAA08036 standard: cDNA, 830 BP.

AC AAA08036;  
 DT 19-JUN-2000 (first entry)  
 XX Human protein transport molecule (PTAM) encoding cDNA SEQ ID NO:10.  
 DE XX

KW Human: protein transport molecule; PTAM; diagnosis; cytostatic;  
 KW antihypertensive; antidiabetic; immunosuppressant; antihypertensive;  
 KW antiallergic; antidiabetic; antipsychotic; antihypertensive; antipsychotic;  
 KW dermatological; antidiabetic; antipsychotic; antihypertensive; antipsychotic;  
 KW immunofluorescent; antihypertensive; antipsychotic; antihypertensive; antipsychotic;  
 KW immune disorder; cell proliferative disorder; secretory disorder;  
 KW urticaria; allergy; abnormal vesicle trafficking; asthma;  
 KW autoimmune haemolytic anaemia; ss.

OS Homo sapiens.

PN WO200012703-A2.

PD 09-MAR-2000.

PF 26-AUG-1999; 99WO-US19616.

PR 27-AUG-1998; 98US-0098206.

PA (INCY-) INCYTE PHARM INC.

PI Tang YF, Lai P, Bandman O, Yue H, Corley NC, Guegler KJ;  
 PI Gorgone GA, Baughn MR, Patterson C;

DR WPI: 2000-256642/22.

P-PSDB: AAY82318.

PT New human protein transport-associated polypeptide and polynucleotide  
 PT useful for diagnosis, prevention and treatment of cell proliferative  
 PT and secretory disorders such as leukemia, cystic fibrosis  
 PS Claim 9; Page 70; 75pp; English.

CC AAA08035 to AAA08042 encode the human protein transport-associated  
 CC molecules (PTAMs) given in AAY82317 to AAY82324. The PTAMs have  
 CC cytosolic, antihypertensive, antidiabetic, immunosuppressant,  
 CC antihypertensive, antidiabetic, antipsychotic, antihypertensive,  
 CC antihypertensive, antidiabetic, antipsychotic, antihypertensive,  
 CC hepatotropic, antipsychotic, dermatological, antineoplastic, antipsychotic,  
 CC regulator protein transport, PTAM proteins and antagonists are useful for  
 CC preventing or treating a disorder associated with decreased or increased  
 CC expression or activity of PTAM. PTAM polynucleotides are useful for  
 CC diagnosing conditions associated with PTAM, comprising detecting PTAM by  
 CC forming a hybridisation complex, preferably after PCR amplifying the  
 CC biological sample. Diseases prevented, treated or diagnosed include cell  
 CC proliferative disorders such as cancers, immune disorders, secretory  
 CC disorders and other conditions associated with abnormal vesicle  
 CC trafficking, such as allergies, asthma, urticaria and autoimmune  
 CC haemolytic anaemia. Anti-PTAM antibodies may be used as antagonists, as  
 CC a targeting or delivery mechanism for bringing pharmaceutical agents  
 CC into cells or tissues expressing PTAM and for diagnosis of PTAM-related  
 CC disorders. PTAM, its catalytic or immunogenic fragments are useful for  
 CC drug screening using libraries of compounds. PTAM polynucleotides are  
 CC useful for generating hybridisation probes useful in mapping the  
 CC naturally occurring genomic sequences.

SQ Sequence 830 BP; 198 A; 208 C; 195 G; 229 T; 0 other:

Query Match 22.0%; Score 251.6; DB 21; Length 830;

Best Local Similarity 72.4%; Pred. No. 4e-56;  
 Matches 326; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 162 COTCTCTGATTTTAACTTATGTAGATCAGGATGATGAGCTCTGAGAGATTGCA 221  
 DB 114 CATCTGTGATTTTAAAGACCTATGTGATGATGAGCTCTGAGAGATTGCA 173  
 QY 222 ATATTCTATGAGCAATGATTAAGAGAGCGGCACTAACAGCTGTATCTGACA 281  
 DB 174 ATGTCTACTACCAACCATGATGATGAGCGGCGCTTCTGCTCCCTGTACAGAGGCA 233  
 QY 282 AGGCCACCTTAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 341  
 DB 234 CAGCCACCTGTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 293

OY		342	TTGACACATTCCTCCTGTAGTGCAGTGCAATATGTAGATTGCCAACCGATTGATG	402
Db		294	TTCGAATGTGGCTTCCACCGAGTTCCAAATACGCGTGGTAGCTCCACGCTTTGATG	353
OY		402	AGCAACGACATCATGCCAAMAATAGATGTTGTTGTGACCAGTGGAACTGTGAATTTG	461
Db		354	ATGAAGCCACACCAAGCCAGACACGAGTCTGTTGTATCTGTGTGATAGTAAGTTTG	413
OY		462	ATGGAACAACACACATTTTCTTCAACACAGACTCTTCCTGTACTGCTACATGCATCCCA	521
Db		414	AGGGGAAACAAACACAGGCGACTTCAACACGAATCTATCTGACCGCCAGGCTCACCCA	473
OY		522	ACAAATACGTGGGAAAGATGTGAATGATTCGTTCCGTTTTCAGAATTTGGTCTAGTAGTT	581
Db		474	GCAACACAGTGTGGGAAGATGCGAAGTGAAGTCTGCTCCGTTCCAGAGACTGGGCCAGCTAGT	533
OY		582	AAAGGGCAAAAAGTCCATTTCTCATTTGGTGC	611
Db		534	GGGGGTGGCAGAGGCTCTTTCCTTCATTC	563
 RESULT 11 AAS26215 standard; cDNA: 864 BP.				
XX	AC	AAS26215;		
XX	DT	07-NOV-2001 (first entry)		
DE	XX	Human cDNA encoding a novel secreted protein, Seg ID 394.		
KW	KM	Human; immunosuppressive; antiarthritic; ss; antirheumatic; cytostatic; cardiant; vasotropic; cerebroprotective; neurotic; neuroprotective; antibacterial; virucide; fungicide; ophthalmological; vulnerable; secreted protein; rheumatoid arthritis; hyperproliferative disorder; cardiovascular disorder; cardiac arrest; cerebrovascular disorder; cerebral ischemia; angiogenesis; nervous system disorder; Alzheimer's disease; infection; ocular disorder; corneal infection; wound healing; epithelial cell proliferation; skin ageing; food additive; preservative; antiproliferative.		
OS	XX	Homo sapiens.		
PN	FN	M0200155322-AZ.		
XX	PD	02-AUG-2001.		
XX	PF	17-JAN-2001; 2001WO-US01341.		
PR	PR	31-JAN-2000; 2000US-0179065.		
PR	PR	04-FEB-2000; 2000US-0180628.		
PR	PR	24-FEB-2000; 2000US-0184664.		
PR	PR	02-MAR-2000; 2000US-0186350.		
PR	PR	16-MAR-2000; 2000US-0189874.		
PR	PR	17-MAR-2000; 2000US-0190076.		
PR	PR	18-APR-2000; 2000US-0198123.		
PR	PR	19-MAY-2000; 2000US-0205515.		
PR	PR	07-JUN-2000; 2000US-0209467.		
PR	PR	28-JUN-2000; 2000US-0214886.		
PR	PR	30-JUN-2000; 2000US-0215135.		
PR	PR	07-JUL-2000; 2000US-0216647.		
PR	PR	07-JUL-2000; 2000US-0216880.		
PR	PR	11-JUL-2000; 2000US-0217487.		
PR	PR	11-JUL-2000; 2000US-0217496.		
PR	PR	14-JUL-2000; 2000US-0218290.		
PR	PR	26-JUL-2000; 2000US-0220963.		
PR	PR	26-JUL-2000; 2000US-0220964.		
PR	PR	14-AUG-2000; 2000US-0224518.		
PR	PR	14-AUG-2000; 2000US-0224519.		
PR	PR	14-AUG-2000; 2000US-0225213.		
PR	PR	14-AUG-2000; 2000US-0225214.		
PR	PR	14-AUG-2000; 2000US-0225266.		



PI Schlegel R, Endege WO, Monahan JE;  
XX WPI: 2001-662795/76.  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer  
XX  
XX Claim 1: Page 1144-1145; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (1) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (1) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 245 BP; 79 A; 47 C; 46 G; 73 T; 0 other;  
XX  
Query Match 19.2%; Score 220; DB 23; Length 245;  
Best Local Similarity 100.0%; Pred. No. 5.2e-48;  
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 837 GGTACTACATCTCTGCTGTAACCCAGCCCTATTGCAAAAATAATCTTTTATATAT 896  
DB 244 GGTACTACATCTCTGCTGTAACCCAGCCCTATTGCAAAAATAATCTTTTATATAT 185  
QY 897 TATCATGAGATGTCAGCACAATATATACACTCTGGGAAGAAGTGGAGTTTGGTTAT 956  
DB 184 TATCATGAGATGTCAGCACAATATATACACTCTGGGAAGAAGTGGAGTTTGGTTAT 125  
QY 957 AGGTTAATTTTCTAGTAACAAACAGTTCCTGTTTTCAGTTAACACATGTAATGCCATTTT 1016  
DB 124 AGGTTAATTTTCTAGTAACAAACAGTTCCTGTTTTCAGTTAACACATGTAATGCCATTTT 65  
QY 1017 AATATATGGCTTTTCAAAATCAGTTCACTGTAAGAAATAGTAC 1056  
DB 64 AATATATGGCTTTTCAAAATCAGTTCACTGTAAGAAATAGTAC 25  
XX  
RESULT 13  
ABV36989/C  
ID ABV36989 standard; cDNA: 441 BP.  
XX  
XX ABV36989;  
XX  
XX 16-SEP-2002 (first entry)  
XX  
XX Human prostate expression marker cDNA 36980.  
XX  
XX Human; prostate cancer; cytostatic; carcinogen; pharmacodynamic marker;  
XX pharmacogenomic marker; gene; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200160860-A2.  
XX  
XX 23-AUG-2001.  
XX  
XX 20-FEB-2001; 2001WO-US05171.  
XX  
XX 17-FEB-2000; 2000US-183319P.  
XX  
XX 16-MAR-2000; 2000US-189862P.  
XX  
XX 25-MAY-2000; 2000US-207454P.  
XX

PR 09-JUN-2000; 2000US-211314P.  
PR 18-JUL-2000; 2000US-219007P.  
PR 13-DEC-2000; 2000US-255281P.  
XX  
XX (MILL-) MILLENNIUM PREDICTIVE MEDICINE INC.  
XX  
XX Schlegel R, Endege WO, Monahan JE;  
XX WPI: 2001-662795/76.  
XX  
XX Novel isolated nucleic acid molecule associated with cancerous state of  
PT prostate cells and correlating with presence of prostate cancer, useful  
PT for detecting presence of prostate cancer, stage of prostate cancer  
XX  
XX Claim 1: Page 7616; 11750pp; English.  
XX  
XX The invention relates to an isolated nucleic acid molecule (1) comprising  
CC a nucleotide sequence given in Tables 1-9 (ABV00010-ABV62213) of the  
CC specification or its complement. (1) is useful for:  
CC (a) assessing whether a patient is afflicted with prostate cancer;  
CC (b) monitoring the progression of prostate cancer in a patient;  
CC (c) assessing the efficacy of a test compound to inhibit prostate  
CC cancer in a patient;  
CC (d) assessing the efficacy of a therapy for inhibiting prostate cancer  
CC in a patient;  
CC (e) selecting a composition for inhibiting prostate cancer in a patient;  
CC (f) assessing the prostate cell carcinogenic potential of a compound;  
CC (g) determining whether prostate cancer has metastasized in a patient;  
CC (h) assessing the aggressiveness or indolence of prostate cancer in a  
CC patient;  
CC (i) is also useful as a pharmacodynamic or pharmacogenomic marker.  
XX  
SQ Sequence 441 BP; 118 A; 95 C; 102 G; 126 T; 0 other;  
XX  
Query Match 19.2%; Score 220; DB 23; Length 441;  
Best Local Similarity 100.0%; Pred. No. 6.4e-48;  
Matches 220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
XX  
QY 837 GGTACTACATCTCTGCTGTAACCCAGCCCTATTGCAAAAATAATCTTTTATATAT 896  
DB 271 GGTACTACATCTCTGCTGTAACCCAGCCCTATTGCAAAAATAATCTTTTATATAT 212  
QY 897 TATCATGAGATGTCAGCACAATATATACACTCTGGGAAGAAGTGGAGTTTGGTTAT 956  
DB 211 TATCATGAGATGTCAGCACAATATATACACTCTGGGAAGAAGTGGAGTTTGGTTAT 152  
QY 957 AGGTTAATTTTCTAGTAACAAACAGTTCCTGTTTTCAGTTAACACATGTAATGCCATTTT 1016  
DB 151 AGGTTAATTTTCTAGTAACAAACAGTTCCTGTTTTCAGTTAACACATGTAATGCCATTTT 92  
QY 1017 AATATATGGCTTTTCAAAATCAGTTCACTGTAAGAAATAGTAC 1056  
DB 91 AATATATGGCTTTTCAAAATCAGTTCACTGTAAGAAATAGTAC 52  
XX  
RESULT 14  
AAS58525  
ID AAS58525 standard; cDNA: 318 BP.  
XX  
XX AAS58525;  
XX  
XX 13-FEB-2002 (first entry)  
XX  
XX cDNA #1201 encoding portion of a human colon tumour protein.  
XX  
XX Human; colon tumour protein; colon cancer; gene therapy; cytostatic; ss.  
XX  
XX Homo sapiens.  
XX  
XX WO200173027-A2.  
XX  
XX 04-OCT-2001.  
XX

```
PF 22-MAR-2001; 2001WO-US09246.
XX
PR 24-MAR-2000; 2000US-191597P.
PR 04-MAY-2000; 2000US-202024P.
PR 05-MAY-2000; 2000US-202189P.
XX
XX (CORI-) CORIXA CORP.
PA
XX
XX Meagher MJ, Xu J, King GE;
PI
XX WPI; 2001-611627/70.
XX
XX New colon tumour proteins and related nucleic acid, useful for
XX treatment, prevention, diagnosis and monitoring of cancer
XX
XX Claim 4; Page 246; 299pp; English.
XX
XX Th present invention relates to the isolation of novel cDNA sequences
XX encoding for at least an immunogenic portion of human colon tumour
XX proteins. The sequences of the invention are useful in pharmaceutical
XX compositions and vaccines for the prevention and treatment of cancers
XX such as colon cancer. They are also useful for the diagnosis and
XX monitoring of such cancers. Antibodies to the colon tumour proteins
XX and antigen presenting cells that express polynucleotides encoding
XX colon tumour proteins can be used to inhibit the development of
XX cancers. T-cells that react specifically with colon tumour proteins
XX are useful for removing tumour cells from samples (e.g. blood) and
XX for cancer treatment. The polynucleotides sequences are also useful in
XX gene therapy. AAS57325-AAS58880 represent the cDNA sequences of the
XX invention that encode for portions of human colon tumour proteins.
XX
XX Sequence 318 BP; 76 A; 82 C; 83 G; 76 T; 1 other;
XX
XX Query Match 16.1%; Score 184.8; DB 23; Length 318;
XX Best Local Similarity 73.8%; Pred. No. 1e-38;
XX Matches 234; Conservative 0; Mismatches 83; Indels 0; Gaps 0;
OY 195 CATGTAGAGCTGCTGAGGAGTTTGTCAATATTACTATGACAAATGATTAAGAAGAC 254
DB 1 CCTGCAGAGCTGCTGAGGAGTTTGTCAATGCTACTACACCACCATGATGAAGCGCGGC 60
OY 255 GGGCACTAACCGGCTGATCTGGACAGCCACCTTAATATGGAATGGAATGCTGTTT 314
DB 61 GTTGCTGCTCCGCTGTACATGGCAGCAGCCACCTGCTGGAATGGCAATGCTGTTT 120
OY 315 CAGGCTGGATGCCCTTAATAATTTTTTACACATTCGCTTCTAGTGAGTTCCAGGTCA 374
DB 121 CAGGACAAAGAAATCCTTGAGTGAGTTTGAATGTTGCTTCCAGCGAGTTCCAAATCA 180
OY 375 ATATGTTAGATGTCGAACCACTTCATGAGCAAGCAACTCACTCCCAAACTACAGTTCTTG 434
DB 181 CGGTGGTAGACTGCCGCTTTCATGATGAAGCCACACACAGCCAGCACCGTCTCTTG 240
OY 435 TTGTGACCAAGTGAAGTTGATGGAACAAACAACTTCTTCAACAGAACT 494
DB 241 TTGTCATCTGGATCAGTGAGTTTGAGGGGACAAACAAACAGGACCTTCAACAGAACT 300
OY 495 TCCTGCTGACTGCTCAG 511
DB 301 TCATNCTGACCGCCAG 317
RESULT 15
ABN38056
ID -ABN38056 standard; DNA; 60 BP.
XX
XX AC ABN38056;
XX
XX 15-JUL-2002 (first entry)
XX Human spliced transcript detection oligonucleotide SEQ ID NO:10804.
XX Human; mouse; rat; splice transcript; detection; RNA transcript;
```

KW: splice variant; transcriptome; oligonucleotide library; ss.

OS Homo sapiens.

PN WO200210449-A2.

XX 07-FEB-2002.

XX 20-JUL-2001; 2001WO-IB01903.

XX 28-JUL-2000; 2000US-221607P.

PR 02-MAY-2001; 2001US-287724P.

XX (COMP-) COMPUGEN INC.

XX Shoshan A, Wasserman A, Mintz E, Mintz L, Faigler S;

PI WPI; 2002-257383/30.

XX New oligonucleotide libraries comprising oligonucleotides which  
XX selectively hybridize to mRNAs transcribed from a transcription unit of  
XX a genome, useful for detecting tissue-, pathology-, and  
XX developmental-specific genes

PS Example 1; SEQ ID 10804; 47pp; English.

XX The present invention describes oligonucleotide libraries for detecting  
XX messenger RNAs that populate a (sub-)transcriptome, where the  
XX (sub-)transcriptome comprises messenger RNAs transcribed from multiple  
XX transcription units that populate a genome. The library comprises  
XX several oligonucleotides, each capable of hybridising selectively to a  
XX set of messenger RNAs transcribed from a given transcription unit of  
XX the genome, which encodes one or more messenger RNA splice variants.  
XX The oligonucleotide libraries are useful for detecting mRNAs from a  
XX biological sample, in expression profiling studies, in qualitatively or  
XX quantitatively characterising the corresponding transcriptome, and in  
XX detecting RNA transcripts and splice variants of human or animal  
XX transcriptomes. The libraries may also be used as specialised mini  
XX libraries to detect transcripts of a sub-transcriptome under a  
XX particular biological or pathological state, and so allowing the  
XX detection of tissue- and pathology-specific genes such as those genes  
XX only expressed in specific tissue under a specific pathological  
XX condition; to detect developmental specific genes; and to detect RNA  
XX transcripts and splice variants of a transcriptome of a patient suffering  
XX from a particular disorder. ABN27253 to ABN59589 represent  
XX oligonucleotide sequences from rats, humans and mice, which are used in  
XX the exemplification of the present invention.  
XX N.B. the sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pct\_sequences.

XX Sequence 60 BP; 17 A; 17 C; 11 G; 15 T; 0 other;

Query Match 5.2%; Score 60; DB 24; Length 60;

Best Local Similarity 100.0%; Pred. No. 3.4e-06;

Matches 60; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 383 GATTGCCAACCAAGTTCATGAGCAAGCAACTCAGTCCCAAACTACAGTTCTTGTGTGACC 442

DB 1 GATTGCCAACCAAGTTCATGAGCAAGCAACTCAGTCCCAAACTACAGTTCTTGTGTGACC 60

Search completed: June 28, 2003, 01:09:33

Job time : 296.199 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2003, 00:22:51 ; Search time 3099.3 seconds

(Without alignments)  
10761.067 Million cell updates/sec

Title: US-09-763-902B-15

Perfect score: 1146

Sequence: 1 tactccgagagagagatggg.....tttaagtaattttgtttt 1146

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : GenEmbl:  
1: gb\_ba:\*  
2: gb\_htg:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pl:\*  
9: gb\_pt:\*  
10: gb\_ro:\*  
11: gb\_sts:\*  
12: gb\_sy:\*  
13: gb\_un:\*  
14: gb\_vl:\*  
15: em\_ba:\*  
16: em\_fun:\*  
17: em\_hum:\*  
18: em\_in:\*  
19: em\_mu:\*  
20: em\_om:\*  
21: em\_ov:\*  
22: em\_ov:\*  
23: em\_pat:\*  
24: em\_ph:\*  
25: em\_pl:\*  
26: em\_ro:\*  
27: em\_sts:\*  
28: em\_un:\*  
29: em\_vl:\*  
30: em\_htg\_hum:\*  
31: em\_htg\_iny:\*  
32: em\_htg\_other:\*  
33: em\_htg\_mus:\*  
34: em\_htg\_pln:\*  
35: em\_htg\_rtd:\*  
36: em\_htg\_rtd:\*  
37: em\_htg\_vrt:\*  
38: em\_sy:\*  
39: em\_htgo\_hum:\*  
40: em\_htgo\_mus:\*  
41: em\_htgo\_other:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1123	98.0	2577	9	BC014888 Homo sapi
2	1064	92.8	2692	9	AK023289 Homo sapi
3	888	77.5	896	9	AF201942 Homo sapi
4	747.2	65.2	178863	9	AF201942 Homo sapi
5	429	37.4	429	9	AF246127 Homo sapi
6	429	37.4	429	9	AF246127 Homo sapi
7	415	36.2	567	9	HS4277591 Homo sapi
8	301.4	26.3	148320	2	AC107438 Homo sapi
9	301.4	26.3	173939	2	AC110357 Rattus no
10	301.4	26.3	186222	2	AC123243 Rattus no
11	274.2	23.9	232171	10	AL731672 Mouse DNA
12	251.6	22.0	939	9	BC002687 Homo sapi
13	251.6	22.0	963	9	BC003029 Homo sapi
14	251.6	22.0	965	9	AF156957 Homo sapi
15	251.6	22.0	972	9	BC000759 Homo sapi
16	251.6	22.0	1176	9	BC003410 Homo sapi
17	251.6	22.0	112180	9	HS1322613 Homo sapi
18	250.6	21.9	1049	9	AK026360 Homo sapi
19	245.8	21.4	186415	2	AC110699 Rattus no
20	243.8	21.3	644	10	AF156958 Mus muscu
21	186.2	16.2	182118	2	AC128310 Rattus no
22	184.8	16.1	318	6	AX261550 Sequence
23	182	15.9	135619	9	AL354809 Human DNA
24	182	15.9	147310	2	AL513495 Homo sapi
25	168.4	14.7	295166	2	AL161914 Homo sapi
26	164.4	14.3	153483	2	AC111510 Rattus no
27	142	12.4	207621	2	AL845495 Mus muscu
28	120.4	10.5	175991	2	AC094606 Rattus no
29	75.6	6.7	203822	2	AC073698 Mus muscu
30	73.4	6.4	173939	2	AC110357 Rattus no
31	61.4	5.4	186222	2	AC123243 Rattus no
32	55.8	4.9	203822	2	AC073698 Mus muscu
33	55.6	4.9	507	3	AF156959 Drosophila
34	55.6	4.9	538	3	AV071250 Drosophila
35	53.2	4.6	1141	6	AX083744 Sequence
36	51.8	4.5	187013	2	AC116920 Dictyostel
37	51.2	4.5	1141	6	AX083744 Sequence
38	50.6	4.4	86827	3	PFMAL3P5 AX083744 Sequence
39	50.6	4.4	349980	6	AX344564 Plasmodu
40	50	4.4	310779	2	AC005140 Plasmodu
41	49.2	4.3	17738	6	AX346441 Sequence
42	48.8	4.3	1055	6	AX083745 Sequence
43	48.8	4.3	66016	2	AC011162 Homo sapi
44	48.2	4.2	29016	2	AC117266 Dictyoste
45	48	4.2	14001	3	PFCOMPIRB X95276 P. falciparu

## ALIGNMENTS

RESULT 1  
LOCUS BC014888 2577 bp mRNA PRI 29-OCT-2001  
DEFINITION Homo sapiens, hypothetical protein P15-2, clone MGC:10178  
IMAGE:3908026, mRNA, complete cds.  
ACCESSION BC014888  
VERSION BC014888.1 GI:15928849  
KEYWORDS MGC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
Strausberg, R.  
TITLE Direct Submission

Pred. No. is the number of results predicted by chance to have a



Sugawara, M., Takahashi, M., Chiba, Y., Ishida, S., Murakawa, K.,  
 Ono, Y., Takiguchi, S., Watanabe, S., Kimura, K., Murakami, K.,  
 Ishii, S., Kawai, Y., Saito, K., Yamamoto, J., Wakamatsu, A.,  
 Nakamura, Y., Nagahara, K., Masuo, Y., Ninomiya, K. and Iwayanagi, T.  
 NEDO human cDNA sequencing project  
 Unpublished  
 TITLE  
 2 (bases 1 to 2692)  
 Isogai, T. and Otsuki, T.  
 Direct Submission  
 Submitted (23-AUG-2000) Takao Isogai, Helix Research Institute,  
 Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan  
 (E-mail: genomics@helix.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)  
 NEDO human cDNA sequencing project supported by Ministry of  
 International Trade and Industry of Japan: cDNA full insert  
 sequencing: Research Association for Biotechnology: cDNA library,  
 construction, 5'- & 3'-end one pass sequencing and clone selection:  
 Helix Research Institute (supported by Japan Key Technology Center  
 etc.) and Department of Virology, Institute of Medical Science,  
 University of Tokyo.  
 Location/Qualifiers  
 1..2692  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="OVARC1000071"  
 /tissue\_type="ovary, tumor tissue"  
 /clone\_1id="OVARC1"  
 /note="Cloning vector: pME18SFL3"  
 103..696  
 /note="unnamed protein product"  
 /codon\_start=1  
 /protein\_id="BAB14511.1"  
 /db\_xref="GI:10435161"  
 /translation="MKYRSHWSOGREGYORRSHNYEGPHSHSSPADRTREEVTP  
 LPEHATRSQMATSLDEFTYDQACRAAEFVNIYEEMDKRRALTRYLDKATLI  
 NMGNAVSGDLNMFEDLPSEFEQVMMDQVHEQAOQSOTTVLVSGTVKPFQGN  
 KQHPNOMNLNLAOSPPNPTWKIASDCRFQDMSS"  
 BASE COUNT 853 a 471 c 478 g 890 t.  
 ORIGIN  
 Query Match 92.8%; Score 1064; DB 9; Length 2692;  
 Best Local Similarity 100.0%; Pred. No. 1.1e-230;  
 Matches 1064; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 616 ACTGCTACAGTCACCTCCCAACAATAGTGTGGAGATTGCAAGTATGCTCCGTTT 675  
 QY 563 CAAGATTGGCTGATGTTAAAGGGCAAAAGTCCATCTTCATTGGTCCATTAGTCCA 622  
 Db 676 CAAGATTGGCTGATGTTAAAGGGCAAAAGTCCATCTTCATTGGTCCATTAGTCCA 735  
 QY 623 GCAATTGGAATTAATGTAATTTTGGATGTAGAGAGCACTTAATPAGTGTGAAC 662  
 Db 736 GCATTTGAATTAATGTAATTTTGGATGTAGAGAGCACTTAATPAGTGTGAAC 795  
 QY 683 AATTTCTTAATTAATTTTCTATCTGTCAGACCTTTCTACACCTGCGACATTGGAG 742  
 Db 796 AATTTCTTAATTAATTTTCTATCTGTCAGACCTTTCTACACCTGCGACATTGGAG 855  
 QY 743 CATGGCCCTTAAGACCTTAAACATATTTTACATGCTTATATACATTCACATA 802  
 Db 856 CATGGCCCTTAAGACCTTAAACATATTTTACATGCTTATATACATTCACATA 915  
 QY 803 GACATCTTAATTAATTAATTAACATATCTTGTACTAATCTCAGTGAACCC 862  
 Db 916 GACATCTTAATTAATTAATTAACATATCTTGTACTAATCTCAGTGAACCC 975  
 QY 863 AGCCTATTCGCAAAATTAATCTTTTATATATATATATATATGATGTCAGCAATATA 922  
 Db 976 AGCCTATTCGCAAAATTAATCTTTTATATATATATATATATGATGTCAGCAATATA 1035  
 QY 923 ACACCTGGGAGAGAGAGATTTTGGTATTAATTAATTTCTAGTAAACAGCTT 982  
 Db 1036 ACACCTGGGAGAGAGAGATTTTGGTATTAATTAATTTCTAGTAAACAGCTT 1095  
 QY 983 GCGTGTTCAGTTAACTGCTGTAATGCCATTTTATATATATGCTTTTCAATCAGTTC 1042  
 Db 1096 GCGTGTTCAGTTAACTGCTGTAATGCCATTTTATATATATGCTTTTCAATCAGTTC 1155  
 QY 1043 AGCGAAATATGTCAGATTAAGTTAACAATCTGACATCTGACATCTGACATCTGACAT 1102  
 Db 1156 AGCGAAATATGTCAGATTAAGTTAACAATCTGACATCTGACATCTGACATCTGACAT 1215  
 QY 1103 GAGATGTCAGTGTGTTTTCATTTAATTAATTTTGT 1146  
 Db 1216 GAGATGTCAGTGTGTTTTCATTTAATTAATTTTGT 1259  
 QY  
 Db  
 RESULT 3  
 AF201942 896 bp mRNA linear PRI 20-JUL-2000  
 LOCUS  
 DEFINITION Homo sapiens DC9 (DC9) mRNA, complete cds.  
 ACCESSION AF201942  
 VERSION AF201942.1 GI:9295187  
 KEYWORDS FLI-CDNA.  
 SOURCE  
 ORGANISM Homo sapiens.  
 Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 REFERENCE  
 1 (bases 1 to 896)  
 AUTHORS Li, Y., Peng, Y., Li, N., Gu, W., Han, Z., Fu, G. and Chen, Z.  
 TITLE Novel genes expressed in human dendritic cell  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 896)  
 AUTHORS Li, Y., Peng, Y., Li, N., Gu, W., Han, Z., Fu, G. and Chen, Z.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-NOV-1999) Chinese National Human Genome Center at  
 Shanghai, 351 Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong,  
 Shanghai 201203, P. R. China  
 Location/Qualifiers  
 1..896  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /cell\_type="dendritic cell"  
 1..896  
 /gene="DC9"  
 145..573  
 /gene="DC9"  
 CDS  
 gene  
 source

```

/codon_start=1
/product="DC9"
/protein_id="AA086878.1"
/db_xref="GI:9295188"
/translation="MATSLEPKTYVDQACRAAEFVNIYIEYMDKRRRLRLYLKDA
TLINGNVAVSLDALNFDLTPSESEQVNMDCQPVHEQATOSQTVLVTSQYKF
DGNKHFFNFLLTAOSTPNNFTWKIASCFRFDQSSS"
BASE COUNT      267 a      177 c      178 g      274 t
ORIGIN

```

```

Query Match      77.5% Score 888; DB 9; Length 896;
Best Local Similarity 100.0%; Aligned No. 9,2e-191;
Matches 888; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 11 AGAATAAGGAGGATGGAATAATTTTGGCGTTGGCGGTTTCCCTCTCTCATTAAGTAT 70
DB 1 AGAATAAGGAGGATGGAATAATTTTGGCGTTGGCGGTTTCCCTCTCTCATTAAGTAT 60
OY 71 TGATCATTCGCGACCGCTGGGACCGGACACGTGAGAGAGTGAAGCGGACGCTGCCA 130
DB 61 TGATCATTCGCGACCGCTGGGACCGGACACGTGAGAGAGTGAAGCGGACGCTGCCA 120
OY 131 GAACACACTGCTACAGAGTCCAGATGCGCAGCTCTCTGATTTTAAACTTATGATGAT 190
DB 121 GAACACACTGCTACAGAGTCCAGATGCGCAGCTCTCTGATTTTAAACTTATGATGAT 180
OY 191 CAGCATGTAGAGTGTCTGAGAGATTTGTCAATTTTACTATGAGACATGATGATAAGA 250
DB 181 CAGCATGTAGAGTGTCTGAGAGATTTGTCAATTTTACTATGAGACATGATGATAAGA 240
OY 251 AGACGGGACACTAACAGGCTGTATCTGGACCAAGCCACCTTAATGATGATGAAATGCT 310
DB 241 AGACGGGACACTAACAGGCTGTATCTGGACCAAGCCACCTTAATGATGATGAAATGCT 300
OY 311 GTTTCAGGCGTGATGCCCTTAATTAATTTTGTGACACATTCCTTCTAGTGTCCAG 370
DB 301 GTTTCAGGCGTGATGCCCTTAATTAATTTTGTGACACATTCCTTCTAGTGTCCAG 360
OY 371 GTCAATATGTATAGTATGCAACCAAGTATATAGCAAGCACTAGTCCCAACTACAGT 430
DB 361 GTCAATATGTATAGTATGCAACCAAGTATATAGCAAGCACTAGTCCCAACTACAGT 420
OY 431 CTGTGTGACACAGTGAAGTGTGATGATGGAAGCAACAACATTTTCTCAACAG 490
DB 421 CTGTGTGACACAGTGAAGTGTGATGATGGAAGCAACAACATTTTCTCAACAG 480
OY 491 AACTTCCTGCTAGTGTCTAGTCCACTCCCAACATATCTGTGTGAAGATGCAAGTAT 550
DB 481 AACTTCCTGCTAGTGTCTAGTCCACTCCCAACATATCTGTGTGAAGATGCAAGTAT 540
OY 551 TGCCTCCGTTTCAAGATGTGTAGTAAAGGGGCAAAAGTCCATTCATTTGGT 610
DB 541 TGCCTCCGTTTCAAGATGTGTAGTAAAGGGGCAAAAGTCCATTCATTTGGT 600
OY 611 CCATTAGTTCAGCAATGGAATTTATGTAATTTTGTGATGAGAGACATATATA 670
DB 601 CCATTAGTTCAGCAATGGAATTTATGTAATTTTGTGATGAGAGACATATATA 660
OY 671 TGTGCTGAAGTAATTTCTTAATTTTCTATTCCTGTCAGACACCTTTCTAGCAGCT 730
DB 661 TGTGCTGAAGTAATTTCTTAATTTTCTATTCCTGTCAGACACCTTTCTAGCAGCT 720
OY 731 GCCAGTTTGAGCATGTGCTCTAAGAGCTTAAACATATTTTAACTGCTTATATA 790
DB 721 GCCAGTTTGAGCATGTGCTCTAAGAGCTTAAACATATTTTAACTGCTTATATA 780
OY 791 CATTCCTAATGACATCTATATATATATATTAACACATGATCTGTGCTACTACATCT 850
DB 781 CATTCCTAATGACATCTATATATATATATTAACACATGATCTGTGCTACTACATCT 840
OY 851 CACTGTGAAGCCAGCCTATGCAAAAATTAATCTTTTATATATATA 898
DB 841 CACTGTGAAGCCAGCCTATGCAAAAATTAATCTTTTATATATATA 888

```

```

RESULT 4
HS596C15/C
LOCUS
DEFINITION
Human DNA sequence from clone 596C15 on chromosome Xq23. Contains
the GUCY2F gene for guanylate cyclase 2F, retinal (RC 4,6,1,2,
RENGC-2, Rod Outer Segment Membrane Guanylate Cyclase 2, ROS-GC2,
GC-F) and a potentially alternatively spliced novel gene. Contains
ESTs, an STS and GSSs, complete sequence.
ACCESSION
AL031387
VERSION
GI:4493578
KEYWORDS
HNG; GC-F; GUCY2F; RENGC-2; retinal guanylate cyclase 2F; Rod Outer
Segment Membrane Guanylate Cyclase; ROS-GC2.
SOURCE
Homo sapiens.
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 178863)
REFERENCE
Howden, P.
Direct Submission
Submitted (01-APR-1999) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Mar 24, 1999 this sequence version replaced gi:4464241.
During sequence assembly data is compared from overlapping clones.
Where differences are found these are annotated as variations
together with a note of the overlapping clone name. Note that the
variation annotation may not be found in the sequence submission
corresponding to the overlapping clone, as we submit sequences with
only a small overlap as described above.
This sequence is the entire insert of clone 596C15. This sequence
has been finished according to sequence map criteria as follows. An
attempt is made to resolve all sequencing problems, such as
compressions and repeats, but not necessarily within known
annotated human repeat sequence elements (e.g. Alu). Where the
sequence is ambiguous, there is an annotation using the 'unsure'
feature key.
This sequence was generated from part of bacterial clone contigs of
human chromosome X, constructed by the Sanger Centre Chromosome X
Mapping Group. Further information can be found at
http://www.sanger.ac.uk/HGP/ChrX
596C15 is from the library RPC14 constructed at the Roswell Park
Cancer Institute by the group of Pieter de Jong. For further
details see http://bacpac.med.buffalo.edu/VECTOR:pcrpac2.
FEATURES
Source
1..178863
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="X"
/map="q23"
/clone="RP4-596C15"
/clone_11b="RPC1-4"
519..560
/note="21 copies 2 mer ac 95% conserved"
1079..1353
/note="match: GSS AQ132738"
1626..1893
/note="AluB3 repeat: matches 1..283 of consensus"
1963..3662
/note="L1P45 repeat: matches 4450..6143 of consensus"
3683..4070
/note="L1P repeat: matches 2885..3261 of consensus"
complement(4444..12256)
/gene="dJ596C15.1"
complement(join(4444..6630,7518..7662,11006..11092,
11867..>11941))
/gene="dJ596C15.1"
/product="dJ596C15.1.1 (novel protein) (isoform 1)"
/note="match: ESTs AA233523 AA354023 AA333479 AA400543
N27567 AA831754 AA459145 AA328722 R67870 H00878 AA148904
AA26578 N20297 AA782964 AA400435 AA458927 AA526010
AI268152 AA234442 AA148905 AI066541 AI467944 R67871 R25673

```



Db 6517 ACTCCCAACATACCTGCTGGAGATTGCAAGTGATGCTCCGTTTCAAGATTGGTCT 6458  
QY 575 AGTAGTTAAAGGGGCAAAAGTCCATCTCTCATTTGGTCCATTAGTTCACAGCAATGAAT 634  
Db 6457 AGTAGTTAAAGGGGCAAAAGTCCATCTCTCATTTGGTCCATTAGTTCACAGCAATGAAT 6398  
QY 635 TAGTGAATTTATTTGATGTAGAGCACTAATAATATGCTGCAAACTAAATTTCTTTAA 694  
Db 6397 TAGTGAATTTATTTGATGTAGAGCACTAATAATATGCTGCAAACTAAATTTCTTTAA 6338  
QY 695 TATTTTCTATCTCTGTCAGCACCTTTCTGAGCAGCTGCCAGTTTGAGCATTGCCCTCTA 754  
Db 6337 TATTTTCTATCTCTGTCAGCACCTTTCTGAGCAGCTGCCAGTTTGAGCATTGCCCTCTA 6278  
QY 755 AGAGCTTTAAACTATTTTACATGCTTATATACATTTCCACTAATGACATTTCTTTATA 814  
Db 6277 AGAGCTTTAAACTATTTTACATGCTTATATACATTTCCACTAATGACATTTCTTTATA 6218  
QY 815 ATAATATTAACACATGATCTTGGTACTAATACATCTACTGTGAACCCAGCCTATTGCAA 874  
Db 6217 ATAATATTAACACATGATCTTGGTACTAATACATCTACTGTGAACCCAGCCTATTGCAA 6158  
QY 875 AAATAAAATCTTTTATATATATCTATGATGATGTCAGCACAAATATACACTCTGGGAA 934  
Db 6157 AAATAAAATCTTTTATATATATCTATGATGATGTCAGCACAAATATACACTCTGGGAA 6098  
QY 935 GAAGTGGAGTTTTTGGTTATAGTTAATTTCTAGTAAACACAGTTGCCCTGTTTTCAG 994  
Db 6097 GAAGTGGAGTTTTTGGTTATAGTTAATTTCTAGTAAACACAGTTGCCCTGTTTTCAG 6038  
QY 995 TTAACACTGGTAAATGCCATTTAATATATGCGTTTTTCAAACTAGTTCAGTGAATAAT 1054  
Db 6037 TTAACACTGGTAAATGCCATTTAATATATGCGTTTTTCAAACTAGTTCAGTGAATAAT 5978  
QY 1055 ACAGATTTAGTTTACATACTACTGACATCTGCACTGGAATTCATATAGATGTTCACT 1114  
Db 5977 ACAGATTTAGTTTACATACTACTGACATCTGCACTGGAATTCATATAGATGTTCACT 5918  
QY 1115 GGTGCTTTTTCATTTTAAAGTAATTTTGTGTTT 1146  
Db 5917 GGTGCTTTTTCATTTTAAAGTAATTTTGTGTTT 5886

RESULT 5  
AF246127  
LOCUS Homo sapiens P15-2 mRNA, complete cds. PRI 21-JUN-2000  
DEFINITION AF246127  
ACCESSION AF246127.1 GI:8575519  
VERSION AF246127.1  
KEYWORDS  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Kang, Y. and Cullen, B.  
TITLE p15-2, a homologous protein of p15, interacts with Tap  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 429)  
AUTHORS Kang, Y. and Cullen, B.  
TITLE Direct Submission  
JOURNAL Submitted (15-MAR-2000) Genetics, Duke University, 427 CARL  
Building, DMC 3025, Durham, NC 27705, USA  
FEATURES  
source  
1. 429  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
1. 429  
/codon\_start=1  
/product="P15-2"  
/protein\_id="AA078034.1"  
/db\_xref="GI:8575520"  
/translation="MATSLDFKTYVDQACRAAEFVNIYYETMDKRRRLRLYLDKA

TLIWNNGVSLDALNNEFDLTPSEFOVNMDCOPVHEQATQSTQTLVLVVTSGTVKE  
DGNKOHFNQNFLLTAQSTPNNTWKIASDCERFQDWSS"  
BASE COUNT 125 a 85 c 94 g 125 t  
ORIGIN  
Query Match 37.4%; Score 429; DB 9; Length 429;  
Best Local Similarity 100.0%; Pred. No. 8.9e-87;  
Matches 429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 155 ATGGCCACGCTCTCTGGATTTTAAACTTATGATAGATCAGGCATGTAGAGCTCTCTGAGGAG 214  
Db 1 ATGGCCACGCTCTCTGGATTTTAAACTTATGATAGATCAGGCATGTAGAGCTCTCTGAGGAG 60  
QY 215 TTTGTCAATATTATGATAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 274  
Db 61 TTTGTCAATATTATGATAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 120  
QY 275 CTGGACAAGCCACCTTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 334  
Db 121 CTGGACAAGCCACCTTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 180  
QY 335 AATTTTTTGGACAATTTGCTTCTAGTGGATGATGATGATGATGATGATGATGATGATGAT 394  
Db 181 AATTTTTTGGACAATTTGCTTCTAGTGGATGATGATGATGATGATGATGATGATGATGAT 240  
QY 395 GTTCATGAGCAAGCAACTCAGTCCCAAACTACAGTTCTTTGTTGTGACCACTGGAAGTGTG 454  
Db 241 GTTCATGAGCAAGCAACTCAGTCCCAAACTACAGTTCTTTGTTGTGACCACTGGAAGTGTG 300  
QY 455 AGTTTGTGAGCAAGCAACTCAGTCCCAAACTACAGTTCTTTGTTGTGACCACTGGAAGTGTG 514  
Db 301 AGTTTGTGAGCAAGCAACTCAGTCCCAAACTACAGTTCTTTGTTGTGACCACTGGAAGTGTG 360  
QY 515 ACTCCCAACAACTACTGCTGGAGATTGCAAGTATGCTGCTTCCGTTTCAAGATTGGTCT 574  
Db 361 ACTCCCAACAACTACTGCTGGAGATTGCAAGTATGCTGCTTCCGTTTCAAGATTGGTCT 420  
QY 575 AGTAGTTAA 583  
Db 421 AGTAGTTAA 429

RESULT 6  
HSA277591  
LOCUS Homo sapiens mRNA for p15-2a protein (p15-2 gene). PRI 19-NOV-2000  
DEFINITION HSA277591  
ACCESSION AJ277591  
VERSION AJ277591.1 GI:8920231  
KEYWORDS p15-2 gene; p15-2a protein.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
AUTHORS Herold, A., Suyama, M., Rodrigues, J.P., Braun, I.C., Kutay, U.,  
Carmona-Fonseca, C., Bork, P. and Izaurralde, E.  
TITLE TAP (NXF1) belongs to a multigene family of putative RNA export  
factors with a conserved modular architecture  
JOURNAL Mol. Cell. Biol. 23, 8996-9008 (2000)  
REFERENCE 2 (bases 1 to 429)  
AUTHORS Izaurralde, E.  
TITLE Direct Submission  
JOURNAL Submitted (26-APR-2000) Izaurralde E., Gene Expression, EMBL,  
Meyerhofstrasse 1, 69117 Heidelberg, GERMANY  
COMMENT Related sequence: AJ278323  
FEATURES  
source  
Location/Qualifiers  
1. 429  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="Xq23"  
/tissue\_type="testis"  
1. 429  
gene

## CDS

/gene="p15-2"  
1.429  
/gene="p15-2"  
/function="role in mRNA nuclear export"  
/note="the p15-2 gene produces two alternatively spliced  
mRNAs encoding for p15-2a and p15-2b proteins"  
/codon\_start=1  
/evidence=experimental  
/product="p15-2a protein"  
/protein\_id="CA8637.1"  
/db\_xref="GI:8920232"  
/translation="MATSLDFKTYVDACRAAEFVNIYETMDKRRRLRLYLKRA  
TLINGNVAGSLDNLNFEFDLPSEFQVNMDCOPHECATOSQTIVLVTSVTK  
DGNKHFNFNLTAAOSTPNNVTWKIASDCFRFDMSSS"  
BASE COUNT 125 a 85 c 94 g 125 t  
ORIGIN

Query Match 37.4%; Score 429; DB 9; Length 429;  
Best Local Similarity 100.0%; Pred. No. 8.9e-87;  
Matches 429; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 155 ATGGCCACGCTCTCGATTTTAAACCTTATGATCAGCATGTAGCTGCTGAGAG 214  
DB 1 ATGGCCACGCTCTCGATTTTAAACCTTATGATCAGCATGTAGCTGCTGAGAG 60  
QY 215 TTGTCAATATTTACTATGAGACATGATTAAGAGAGCGGCACTAACCAAGGCTGTAT 274  
DB 61 TTGTCAATATTTACTATGAGACATGATTAAGAGAGCGGCACTAACCAAGGCTGTAT 120  
QY 275 CTGACACAGGCGACCTTAATATGAAATGAAATGCTTTTCAGGCGCTGATGCCCTAAT 334  
DB 121 CTGACACAGGCGACCTTAATATGAAATGAAATGCTTTTCAGGCGCTGATGCCCTAAT 180  
QY 335 AATTTTGTGACATATGCTCTTCTAGAGAGTCCAGCTCATATGATTGCCAACCA 394  
DB 181 AATTTTGTGACATATGCTCTTCTAGAGAGTCCAGCTCATATGATTGCCAACCA 240  
QY 395 GTTCATGAGCAGCACTAGCTGCTTCTAGAGAGTCCAGCTCATATGATTGCCAACCA 454  
DB 241 GTTCATGAGCAGCACTAGCTGCTTCTAGAGAGTCCAGCTCATATGATTGCCAACCA 300  
QY 455 AAGTTTATGGAACCAACCAACATTTCTCAACAGAGACTTCTGCTAGCTCAGTCC 514  
DB 301 AAGTTTATGGAACCAACCAACATTTCTCAACAGAGACTTCTGCTAGCTCAGTCC 360  
QY 515 ACTGCCACAACTACTGTGTGAGAAATGCAAGTATGCTTCCGTTTCAAGATTGCTCT 574  
DB 361 ACTGCCACAACTACTGTGTGAGAAATGCAAGTATGCTTCCGTTTCAAGATTGCTCT 420  
QY 575 AGTAGTTAA 583  
DB 421 AGTAGTTAA 429

RESULT 7  
LOCUS HSA278323 567 bp mRNA linear PRI 19-NOV-2000  
DEFINITION Homo sapiens mRNA for p15-2b protein (p15-2 gene).  
ACCESSION AJ278323.1 GI:9663146  
VERSION p15-2 gene; p15-2b protein.  
KEYWORDS human.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
REFERENCE 1 (bases 1 to 567)  
AUTHORS Harold, A., Suyama, M., Rodrigues, J.P., Braun, I.C., Kutay, U.,  
Carmona-Fonseca, C., Bork, P. and Izaurralde, E.  
TITLE TAP (NHEJ) belongs to a multigene family of putative RNA export  
factors with a conserved modular architecture  
JOURNAL Mol. Cell. Biol. 23, 8996-9008 (2003)  
REFERENCE 2 (bases 1 to 567)  
AUTHORS Izaurralde, E.

TITLE Direct Submission  
JOURNAL Submitted (06-JUN-2000) Izaurralde E., Gene Expression, EMBL,  
Meyerhofstrasse 1, 69117 Heidelberg, GERMANY  
COMMENT Related sequence: AJ277591.  
FEATURES  
source Location/Qualifiers  
1..567  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="X"  
/map="Xq23"  
/tissue="testis"  
1..567  
/gene="p15-2"  
1..567  
/gene="p15-2"  
/function="role in mRNA nuclear export"  
/note="the p15-2 gene produces two alternatively spliced  
mRNAs encoding for p15-2a and p15-2b proteins"  
/codon\_start=1  
/evidence=experimental  
/product="p15-2b protein"  
/protein\_id="CAC01129.1"  
/db\_xref="GI:9663147"  
/translation="VTNHPALCTAGRGRPRARAPGPPSLVSSRFLIPRTSMEL  
ARGDVGDFKTYVDACRAAEFVNIYETMDKRRRLRLYLKRAITLINGNVAGSL  
DALNFEPTLSSSEFQVNMDCOPHECATOSQTIVLVTSVTKEDGNKHFNFNLT  
AAOSTPNNVTWKIASDCFRFDMSSS"  
BASE COUNT 146 a 131 c 142 g 148 t  
ORIGIN

Query Match 36.2%; Score 415; DB 9; Length 567;  
Best Local Similarity 100.0%; Pred. No. 1.3e-83;  
Matches 415; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 169 GGATTTTAAACCTTATGATGATCAGCATGTAGAGCTGCTGAGAGTGTGCAATATTTA 228  
DB 153 GGATTTTAAACCTTATGATGATCAGCATGTAGAGCTGCTGAGAGTGTGCAATATTTA 212  
QY 229 CTATGACACAACTGATTAAGAGAGCGGCACTAACCAAGGCTGTATCTGACAAAGCCAC 288  
DB 213 CTATGACACAACTGATTAAGAGAGCGGCACTAACCAAGGCTGTATCTGACAAAGCCAC 272  
QY 289 CTTATATGAAATGAAATGCTGTTTCAAGGCTGATGCCCTAATATTTTGTGACAC 348  
DB 273 CTTATATGAAATGAAATGCTGTTTCAAGGCTGATGCCCTAATATTTTGTGACAC 332  
QY 349 ATGCGCTCTGATGAGTCCAGGTCATATGATGATGATGATGATGATGATGATGATGATG 408  
DB 333 ATGCGCTCTGATGAGTCCAGGTCATATGATGATGATGATGATGATGATGATGATGATG 392  
QY 409 AACTCAGTCCCAAACTACAGTCTTGTGTGACCAAGTGAAGTGAAGTGAAGTGAAGTGA 468  
DB 393 AACTCAGTCCCAAACTACAGTCTTGTGTGACCAAGTGAAGTGAAGTGAAGTGAAGTGA 452  
QY 469 CAACACACATTTCTTCAACCAAGACTTCTCTGACAGTCTGACAGTCTGACAGTCTG 528  
DB 453 CAACACACATTTCTTCAACCAAGACTTCTCTGACAGTCTGACAGTCTGACAGTCTG 512  
QY 529 TGTGTGAGAGATGCAAGTGTGCTTCCGTTTCAAGATGGTCTAGTATTAA 583  
DB 513 TGTGTGAGAGATGCAAGTGTGCTTCCGTTTCAAGATGGTCTAGTATTAA 567

RESULT 8  
LOCUS AC107438 148320 bp DNA linear HTG 13-JUN-2002  
DEFINITION Rattus norvegicus clone CH230-160A14, \*\*\* SEQUENCING IN PROGRESS  
ACCESSION AC107438  
VERSION \*\*\* 62 unordered pieces.  
KEYWORDS HTG; HTGS; PHASE1.  
SOURCE Norway rat.  
ORGANISM Rattus norvegicus



REFERENCE  
AUTHORS

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
Rattus.

1 (bases 1 to 148320)

Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-Osman,F.R., Allen,C.,  
Alsbrooks,S.L., Amarantunga,H.C., Are,J.R., Ayele,M., Banks,T.,  
Barbata,J., Benton,J., Blmage,K., Blankenburg,K., Bonini,D.,  
Bouck,J., Bowls,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
Burch,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
Carron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C.,  
Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
Dayla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,  
Delaney,K.R., Delgado,O., Dem,A.L., Ding,Y., Din,H.H.,  
Douthett,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J.,  
Earhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
Falls,T., Ferraguto,D., Flagg,N., Ford,T., Foster,P., Frazz,P.,  
Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
Gorrell,J.H., Guevara,W., Gunaratne,P., Hale,S., Hamilton,K.,  
Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,  
Hernandez,O., Hodgson,A., Hogue,M., Holloway,C., Hollins,B.,  
Horn,J.F., Howard,S., Huber,J., Hulik,S., Hume,J., Jackson,L.E.,  
Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudan,S.,  
Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,  
Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
Li,J., Li,Z., Lichtarge,O., Liu,C., Liu,J., Liu,W., Lousaged,H.,  
Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
Maheshwari,M., Mapa,P., Martin,R., Martindale,A., Martinez,E.,  
Massey,E., Mawhney,E., McLeod,M.P., Meador,M., Mei,G., Metzger,M.,  
Miner,G., Miner,Z., Mitchell,T., Mohabath,K., Morgan,M., Morris,S.,  
Mosier,M., Neal,D., Newton,E., Newton,N., Nguyen,A., Nguyen,N.,  
Nguyen,N., Nicholson,E., Nockemko,S., Ogih,M., Okumou,G.,  
Oragunye,N., Oviedo,R., Pace,A., Payton,B., Peery,J., Perez,L.,  
Peters,L., Pickens,R., Primus,E., Pu,L.L., Quiles,M., Ren,I.,  
Rivers,M., Rojas,A., Rojudoan,I., Rolfe,M., Ruiz,S., Savery,G.,  
Scheer,S., Scott,G., Shen,H., Shoshitari,N., Sisson,I.,  
Sodergren,E., Sonake,T., Sparks,A., Stanley,H., Stone,H.,  
Sutton,A., Swatek,A., Tabor,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
Tansley,D., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
Usmani,R., Vasquez,L., Verra,V., Villalon,D., Vinson,R., Wang,S.,  
Wang,S., Ward-Moore,S., Warren,R., Washington,C., Wallington,S.,  
Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,  
Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
Weinstock,G. and Gibbs,R.

Direct Submission

Unpublished

2 (bases 1 to 148320)

Worley,K.C.

Direct Submission

Submitted (21-JAN-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 148320)

Worley,K.C.

Direct Submission

Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

On Jul 12, 2002 this sequence version replaced gi:18846336.

Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: [hgsc.help@bcm.tmc.edu](mailto:hgsc.help@bcm.tmc.edu)

Project Information

Center Project name: GMZC

Center clone name: CH230-160A14

Summary Statistics

Chemistry: Dye-terminator Big Dye: 100% of reads

Assembly program: Phrap: version 0.990329

Consensus quality: 9845 bases at least Q40

Consensus quality: 94598 bases at least Q30

Consensus quality: 98393 bases at least Q20

\*\*\*\*\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
NOTE: This is a 'working draft' sequence. It currently  
consists of 62 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1	1578:	contig of 1578 bp in length
1579	1678:	gap of unknown length
1679	3151:	contig of 1473 bp in length
3152	3251:	gap of unknown length
3252	4752:	contig of 1501 bp in length
4753	4852:	gap of unknown length
4853	5878:	contig of 1026 bp in length
5879	5978:	gap of unknown length
5979	7250:	contig of 1272 bp in length
7251	7350:	gap of unknown length
7351	8428:	contig of 1078 bp in length
8429	8528:	gap of unknown length
8529	10032:	contig of 1504 bp in length
10033	10132:	gap of unknown length
10133	11243:	contig of 1111 bp in length
11244	11343:	gap of unknown length
11344	12929:	contig of 1586 bp in length
12930	13029:	gap of unknown length
13030	14302:	contig of 1273 bp in length
14303	14402:	gap of unknown length
14403	15468:	contig of 1066 bp in length
15468	15568:	gap of unknown length
15569	17010:	contig of 1442 bp in length
17011	17110:	gap of unknown length
17111	18317:	contig of 1207 bp in length
18318	18417:	gap of unknown length
18418	19702:	contig of 1285 bp in length
19703	19802:	gap of unknown length
19803	21360:	contig of 1558 bp in length
21361	21460:	gap of unknown length
21461	23143:	contig of 1683 bp in length
23144	23243:	gap of unknown length
23244	24991:	contig of 1748 bp in length
24992	25091:	gap of unknown length
25092	25564:	contig of 1473 bp in length
25565	26664:	gap of unknown length
26665	27783:	contig of 1119 bp in length
27784	27883:	gap of unknown length
27884	29648:	contig of 1765 bp in length
29649	29748:	gap of unknown length
29749	31751:	contig of 2003 bp in length
31752	31851:	gap of unknown length
31852	33098:	contig of 1247 bp in length
33099	33198:	gap of unknown length
33199	34820:	contig of 1622 bp in length
34821	34920:	gap of unknown length
34921	36154:	contig of 1234 bp in length
36155	36254:	gap of unknown length
36255	37964:	contig of 1710 bp in length
37965	38064:	gap of unknown length
38065	39211:	contig of 1147 bp in length
39212	39311:	gap of unknown length
39312	41243:	contig of 1932 bp in length
41244	41343:	gap of unknown length
41344	42978:	contig of 1635 bp in length
42979	43078:	gap of unknown length
43079	44146:	contig of 1068 bp in length
44147	44246:	gap of unknown length
44247	45967:	contig of 1721 bp in length
45968	46067:	gap of unknown length
46068	47796:	contig of 1729 bp in length
47797	47896:	gap of unknown length

Query Match	26.38;	Score 301.4;	DB 2;	Length 148320;
Best Local Similarity	73.28;	Pred. No. 3e-58;		
Matches 532;	Conservative	0;	Mismatches 161;	Indels 34;
			Gaps	10;

Db	38776	CTAATGTAATGTAATATTTTTTTTATTCCTAGACAGACACCTTGGCTAGATTAGCTCCAGTT	38835
QY	738	TGAGAGCATTCGCCCTCTAGAGAGCTTTAAACATAATTTTTTTTACATGCGCTTAT--ATACATTC	795
Db	38836	TGGAATTATTCGCCCT--TAAAGAGCTTTTAAAGATGATTTTTTTTACATGCGCTTATGATCATAATTC	38934
QY	796	CACATAATGACATTCCTTATATATATATTTAAACACATGATCTTGGTACATACATGACCT	855
Db	38895	CACATACATGATACCTCCTAGAGTAATATTTAA--ACATATATCAGTACTAAGATACACACAGC	38952
QY	856	TGAGACAGGCTTATGCAAAATAATATCTTTTTTTTAAATATATATATATATGATGATGACAG	915
Db	38853	TGAGACCAATATCTTTGACACAAATGAGAGCTTTTTTTTGTAAATCTGTATGGAAT--AGCA	39011
QY	916	CAATATATACACTCTGGGAGAGAGTGAAGTATTTTTGGTATTTATGATTAATTTTCTAGTAA	975
Db	39012	CAGATATAGCTCTGTAGAGAGAAATATGTGTTTTTTCAGTATATAGGTAATTTCTCAATCA	39071
QY	976	ACACGTTGCTCTGTTTTTCAGTTAACTGATGTAATGCCATTTTAAATATATGCTTTTCAAA	1035
Db	39072	AGCACATGCTCTTTTCAGATGAGAACATGATGATCAAGTTTAAATAT--AAG	39122
QY	1036	TCAGTTCAGCGAATAATAGTACAGATTTTAGCTTTTACATCTACTCTGACATATCTGGAAT	1095
Db	39123	TATAGATATTTTGAAGTACAGTTCATTTAATTTATTTATCTAGTGTCTGTCATATTTCAAT	39182
QY	1096	GCATATTA	1102
Db	39183	GCATATTA	39189
RESULT 9			
AC110357			
LOCUS			
DEFINITION	Ac110357	173939 bp	DNA
	Rattus norvegicus clone CH230-276M2, *** SEQUENCING IN PROGRESS		
	***, 49 unordered pieces.		
AC110357			
AC110357.3	GI:21740885		
HRG:	HRG3_PHASE1.		
KEYWORDS			
SOURCE	Norway rat.		
ORGANISM	Rattus norvegicus		
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;		
	Rattus.		
REFERENCE			
AUTHORS	1 (bases 1 to 173939)		
	Muzny,D.M., Adams,C., Adio-Oduola,B., Ali-osman,F.R., Allen,C.,		
	Aisbrooks,S.L., Amaralunga,H.C., Are,J.R., Ayala,M., Banks,T.,		
	Barbieri,J., Benton,J., Bimagne,K., Blankenburg,K., Bonnin,D.,		
	Bouch,J., Bowie,S., Brilewa,M., Brown,E., Brown,M., Bryant,N.P.,		
	Byham,T., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.,		
	Carroll,T.F., Carter,M., Cavazos,S.R., Checko,J., Chavez,D.,		
	Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,		
	Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,		
	Devilla,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A.,		
	Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Din,h.H.J.,		
	Doutlaite,K.J., Draper,H., Dugan-Rocha,S., Dupin,K.H.,		
	Einhardt,C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M.,		
	Falls,T., Ferriguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,		
	Gabisl,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,		
	Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,		
	Harris,C., Harris,K., Hart,M., Haylak,P., Hawes,A., Hernandez,J.,		
	Hernandez,O., Hodgson,A., Hogues,M., Holloway,C., Hollins,B.,		
	Homs,J., Howard,S., Huber,J., Huliyk,S., Hume,J., Jackson,L.E.,		
	Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,		
	Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C.,		
	Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,		
	Li,J., Li,Z., Lichtarge,O., Lien,C., Liu,J., Liu,W., Louissead,H.,		
	Lozado,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,		
	Maheshwari,M., Mapua,P., Matlin,R., Martindale,A., Matinez,E.,		
	Massej,E., Maxwell,E., McLeod,M.P., Meador,M., Mei,G., Metzker,M.,		
	Miner,G., Miner,Z., Mitchell,T., Mohabbat,K., Morgan,M., Morris,S.,		
	Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,		
	Nguyen,N., Nickerson,E., Nnoekkwu,S., Ogun,H., Okunomu,G.,		

Oragunye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L.L., Quiles, M., Ren, X.,  
Rives, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
Scherer, S., Scott, G., Shen, H., Shoochitari, N., Sisson, I.,  
Sodergren, E., Sotnik, T., Sparks, A., Stanley, H., Stone, H.,  
Sulton, A., Swatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Teiford, B., Thomas, N., Thomas, S.,  
Uman, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wleczek, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,  
Weinstock, G. and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 173939)  
Worley, K.C.  
Direct Submission  
Submitted (11-FEB-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 173939)  
Worley, K.C.  
Direct Submission  
Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA  
On Jul 12, 2002 this sequence version replaced g1:18847131.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: GRM  
Center clone name: CH230-276M2  
----- Summary Statistics  
Sequencing vector: Plasmid;  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 136387 bases at least Q40  
Consensus quality: 139690 bases at least Q30  
Consensus quality: 142385 bases at least Q20

----- NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/genbank/draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 49 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1030: contig of 1030 bp in length  
1031 1130: gap of unknown length  
1131 2654: contig of 1524 bp in length  
2655 2755: gap of unknown length  
2755 3834: contig of 1079 bp in length  
3834 3934: gap of unknown length  
3934 5317: contig of 1383 bp in length  
5317 5416: gap of unknown length  
5416 6880: contig of 1464 bp in length  
6880 6981: gap of unknown length  
6981 7987: contig of 1007 bp in length  
7987 8088: gap of unknown length  
8088 9455: contig of 1368 bp in length  
9455 9555: gap of unknown length  
9555 10645: contig of 1090 bp in length  
10645 10746: gap of unknown length  
10746 11918: contig of 1173 bp in length  
11918 12018: gap of unknown length  
12018 13599: contig of 1581 bp in length  
13599 13600: gap of unknown length

13700 15310: contig of 1611 bp in length  
15310 15410: gap of unknown length  
15410 17265: contig of 1855 bp in length  
17265 17366: gap of unknown length  
17366 19287: contig of 1821 bp in length  
19287 21646: gap of unknown length  
21646 22514: contig of 2359 bp in length  
22514 23515: gap of unknown length  
23515 26113: contig of 1769 bp in length  
26113 26214: gap of unknown length  
26214 28949: contig of 2499 bp in length  
28949 29050: gap of unknown length  
29050 30793: contig of 2736 bp in length  
30793 30892: gap of unknown length  
30892 33814: contig of 1743 bp in length  
33814 33914: gap of unknown length  
33914 37076: contig of 2821 bp in length  
37076 37176: gap of unknown length  
37176 39759: contig of 3163 bp in length  
39759 39860: gap of unknown length  
39860 42980: contig of 2583 bp in length  
42980 45805: gap of unknown length  
45805 45905: contig of 3020 bp in length  
45905 49602: gap of unknown length  
49602 49701: contig of 3020 bp in length  
49701 51822: gap of unknown length  
51822 54942: contig of 2020 bp in length  
54942 55042: gap of unknown length  
55042 57799: contig of 3120 bp in length  
57799 61347: gap of unknown length  
61347 61448: contig of 2757 bp in length  
61448 65731: gap of unknown length  
65731 65832: contig of 4284 bp in length  
65832 67837: gap of unknown length  
67837 67936: contig of 2005 bp in length  
67936 67937: gap of unknown length  
67937 70267: contig of 2330 bp in length  
70267 70367: gap of unknown length  
70367 74979: contig of 4613 bp in length  
74979 75079: gap of unknown length  
75079 78340: contig of 4613 bp in length  
78340 78441: gap of unknown length  
78441 82207: contig of 3767 bp in length  
82207 82308: gap of unknown length  
82308 85016: contig of 2709 bp in length  
85016 85117: gap of unknown length  
85117 88120: contig of 3004 bp in length  
88120 91247: gap of unknown length  
91247 91248: contig of 3027 bp in length  
91248 91348: gap of unknown length  
91348 94541: contig of 3194 bp in length  
94541 100271: gap of unknown length  
100271 100371: contig of 5630 bp in length  
100371 105497: gap of unknown length  
105497 105597: contig of 5126 bp in length  
105597 111059: gap of unknown length  
111059 11159: contig of 5462 bp in length  
11159 115817: gap of unknown length  
115817 115917: contig of 4658 bp in length  
115917 121770: gap of unknown length  
121770 121870: contig of 5853 bp in length  
121870 127414: gap of unknown length  
127414 127514: contig of 5544 bp in length  
127514 135104: gap of unknown length  
135104 135205: contig of 7590 bp in length  
135205 145776: gap of unknown length  
145776 145876: contig of 10572 bp in length  
145876 145877: gap of unknown length  
145877 154867: contig of 8991 bp in length



Summary Statistics

Sequencing vector: plasmid:  
 Chemistry: Dye-terminator Big Dye; 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 123845 bases at least Q40  
 Consensus quality: 128005 bases at least Q30  
 Consensus quality: 131716 bases at least Q20

NOTE: Estimated insert size may differ from sequence length  
 (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 NOTE: This is a 'working draft' sequence. It currently  
 consists of 65 contigs. The true order of the pieces  
 is not known and their order in this sequence record is  
 arbitrary. Gaps between the contigs are represented as  
 runs of N, but the exact sizes of the gaps are unknown.  
 This record will be updated with the finished sequence  
 as soon as it is available and the accession number will  
 be preserved.

1 1632: contig of 1632 bp in length  
 1633 1732: gap of unknown length  
 1733 3449: contig of 1717 bp in length  
 3450 3549: gap of unknown length  
 3550 4717: contig of 1168 bp in length  
 4718 4817: gap of unknown length  
 4818 5981: contig of 1164 bp in length  
 5982 6081: gap of unknown length  
 6082 7502: contig of 1421 bp in length  
 7503 7602: gap of unknown length  
 7603 9031: contig of 1429 bp in length  
 9032 9131: gap of unknown length  
 9132 10390: contig of 1259 bp in length  
 10391 10490: gap of unknown length  
 10491 11746: contig of 1256 bp in length  
 11747 11846: gap of unknown length  
 11847 13717: contig of 1871 bp in length  
 13718 13817: gap of unknown length  
 13818 15195: contig of 1378 bp in length  
 15196 15295: gap of unknown length  
 15296 16397: contig of 1102 bp in length  
 16398 16497: gap of unknown length  
 16498 17899: contig of 1402 bp in length  
 17900 17999: gap of unknown length  
 18000 19371: contig of 1372 bp in length  
 19372 19471: gap of unknown length  
 19472 20716: contig of 1245 bp in length  
 20717 20816: gap of unknown length  
 20817 22597: contig of 1781 bp in length  
 22598 22697: gap of unknown length  
 22698 24044: contig of 1347 bp in length  
 24045 24144: gap of unknown length  
 24145 25471: contig of 1327 bp in length  
 25472 25571: gap of unknown length  
 25572 26916: contig of 1345 bp in length  
 26917 27016: gap of unknown length  
 27017 28825: contig of 1809 bp in length  
 28826 28925: gap of unknown length  
 28926 30360: contig of 1435 bp in length  
 30361 30460: gap of unknown length  
 30461 31852: contig of 1392 bp in length  
 31853 31952: gap of unknown length  
 31953 33709: contig of 1757 bp in length  
 33710 33809: gap of unknown length  
 33810 35330: contig of 1521 bp in length  
 35331 35430: gap of unknown length  
 35431 36971: contig of 1541 bp in length  
 36972 37071: gap of unknown length  
 37072 39696: contig of 2625 bp in length  
 39697 39796: gap of unknown length  
 39797 41922: contig of 2126 bp in length  
 41923 42022: gap of unknown length  
 42023 43620: contig of 1598 bp in length  
 43621 43720: gap of unknown length  
 43721 45330: contig of 1610 bp in length

45331 45430: gap of unknown length  
 45431 47622: contig of 2192 bp in length  
 47623 47722: gap of unknown length  
 47723 49854: contig of 2132 bp in length  
 49855 49954: gap of unknown length  
 49955 51438: contig of 1484 bp in length  
 51439 51538: gap of unknown length  
 51539 53780: contig of 2242 bp in length  
 53781 53880: gap of unknown length  
 53881 55569: contig of 1669 bp in length  
 55570 55670: gap of unknown length  
 55671 58232: contig of 2553 bp in length  
 58233 58332: gap of unknown length  
 58333 60649: contig of 2317 bp in length  
 60650 60749: gap of unknown length  
 60750 63699: contig of 2850 bp in length  
 63700 65434: gap of unknown length  
 65435 65534: contig of 1735 bp in length  
 65535 68538: gap of unknown length  
 68539 70893: contig of 3004 bp in length  
 70894 70993: gap of unknown length  
 70994 73637: gap of 2644 bp in length  
 73638 73737: gap of unknown length  
 73738 75980: contig of 2243 bp in length  
 75981 76080: gap of unknown length  
 76081 79351: contig of 3271 bp in length  
 79352 79451: gap of unknown length  
 79452 81755: contig of 2304 bp in length  
 81756 81855: gap of unknown length  
 81856 84787: contig of 2932 bp in length  
 84788 84887: gap of unknown length  
 84888 87647: contig of 2760 bp in length  
 87648 87747: gap of unknown length  
 87749 91057: contig of 3310 bp in length  
 91058 91157: gap of unknown length  
 91158 94696: contig of 3539 bp in length  
 94697 94796: gap of unknown length  
 94797 96545: contig of 1749 bp in length  
 96546 96645: gap of unknown length  
 96646 100902: contig of 4257 bp in length  
 100903 101002: gap of unknown length  
 101003 105287: contig of 4285 bp in length  
 105288 105387: gap of unknown length  
 105388 108272: contig of 2885 bp in length  
 108273 108372: gap of unknown length  
 108373 112332: contig of 3860 bp in length  
 112333 112332: gap of unknown length  
 112333 114916: contig of 2584 bp in length  
 114917 115017: gap of unknown length  
 115017 118825: contig of 3809 bp in length

Query Match 26.3%; Score 301.4; DB 2; Length 186222;  
 Best Local Similarity 73.2%; Pred. No. 3e-58;  
 Matches 532; Conservative 0; Mismatches 161; Indels 34; Gaps 10;

395 GTTCATGACCAACCACTGACCCCAACTGACCTGCTGTTGACGAGTGAAGTGG 454  
 148228 GCTAAGAGCAACCTACCAAGTACGACGACGACGCTGCTGACCAAGTGGGCT 148169  
 455 AAGTTGATGGAACAACAACATTTCTTCAACAGAACTTCTGCTGACGCTCAGTC 514  
 148168 AAGTTGATGGAACAACAACACTCTTAAACAGAACTTCTGCTGACGAGAGC 148109  
 515 ACTGCCAACAATCTGTGTGGAAGATTGCAAGGATGCTCGCTTTCAAGTTGCTCT 574  
 148108 ACTCCCTAAGAGCAACGAGTGGGAAGATTGCAAGGATGCTCGCTTTCAAGTTG 148049  
 575 AGTAGTAAAGGGCAAGAGTCCATTCATTTGATGCTATTAGTTCAGCAATTAAT 634  
 148048 AGTATTAAAGGAGCAAAAGCCCGT--TGCAATTGGTCATTAGTTCAGCAAGAGATT 147991

635 TA-----TGGAATTAATTTGATTTGATAGACACTATAA-----TATGCTGAAA 680  
 147990 TAAATGTAATTAATGATTTGTTGTCACAAATGATATAAAGTAACTAGTATGCTGATAA 147931  
 681 CTAATATTTCTTAATA-TTTTCTATCTGTCAGACACTTTTCAG--CAGCTGCCACT 737  
 147930 CTAATGATTAATTAATTTTATTTTATCTACAGACACTTTGCTAGATTAAGTGGCAGTT 147871  
 738 TGGAGCAATGGCCCTAGACCTTTAAACATATTTTATACATGCTTAT--ATACATTC 795  
 147870 TGGATTAATGGCCCT-TAAACCTTTAAAGATAGATTTTAAATGCTTATGCTATCAATTC 147812  
 796 CACTAATGACATTTCTTAATTAATTAATTAACATGATCTGCTACTAATACACACATG 855  
 147811 CACTACATGATCTCTCAAGTAATTTTA--ACATGATCTAGTACTAGATACACATG 147754  
 856 TGAACCCAGCCTATTGCAAAAATAAATCTTTTAAATTAATTAATTAATGATGACCA 915  
 147753 TGAACCCAGCCTATTGCAAAAATAAATCTTTTAAATTAATTAATTAATGATGACCA 147695  
 916 CAATTAACACTCTGGGAG 975  
 147694 CAAGTACCTCTGTAAG 147635  
 976 ACAGCTGCTCTGTTTTCAGTTACAGCTGATGATGATGATGATGATGATGATGATGATG 1035  
 147634 AGCAGATGCTCTGTTTTCAGTTACAGCTGATGATGATGATGATGATGATGATGATG 147584  
 1036 TCAGCTCAGTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1095  
 147583 TATAGATTTTGAAGTCAATTTTAAATTAATTAATTAATTAATTAATTAATTAATTA 147524  
 1096 GCATATA 1102  
 147523 GCATATA 147517

RESULT 11  
 AL731672 232171 bp DNA 1linear ROD 20-AUG-2002  
 LOCUS Mouse DNA sequence from clone RP23-161L11 on chromosome X, complete  
 DEFINITION sequence.  
 ACCESSION AL731672  
 VERSION AL731672.11 GI:22415912  
 KEYWORDS HMG.  
 SOURCE house mouse.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 1 (bases 1 to 232171)  
 Clark, S.  
 Direct Submission  
 Submitted (20-AUG-2002) Wellcome Trust Sanger Institute, Hinxton,  
 Cambridgeshire, CB10 1SA, UK. E-mail enquiries:  
 humquerry@sanger.ac.uk Clone requests: clonerequest@sanger.ac.uk  
 On Aug 21, 2002 this sequence version replaced g1:21955555.  
 COMMENT Genome Center  
 Center: Wellcome Trust Sanger Institute  
 Center code: SC  
 Web site: <http://www.sanger.ac.uk>  
 Contact: humquerry@sanger.ac.uk

During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least

one plasmid subclone or more than one MJ3 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em: EMBL; Sw: SWISSPROT; Tr: TrEMBL; Wp: WormPEP; Information on the WormPEP database can be found at [http://www.sanger.ac.uk/projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/projects/C_elegans/wormpep) RP23-161L11 is from the RP23-Mouse PAC library constructed by the group of Pieter de Jong. For further details see <http://www.choir.org/bacpac/home.htm> VECTOR: pBac3.6.

## FEATURES

source  
 1. 232171  
 /organism="Mus musculus"  
 /db\_xref="taxon:10090"  
 /chromosome="X"  
 /clone="RP23-161L11"  
 /clone\_11b="RP23-161L11"  
 /clone\_11b="RP23-161L11"  
 BASE COUNT 71072 a 46436 c 45947 g 68716 t  
 ORIGIN

Query Match 23.9%; Score 274.2; DB 10; Length 232171;  
 Best Local Similarity 72.6%; Pred. No. 4,1e-52;

Matches 526; Conservative 0; Mismatches 143; Indels 56; Gaps 11;

395 GTTCATGAGCAAGCAACATCAGTCCCAATCAGTCTGTTGTCAGCAGTGGAACTGTG 454  
 45590 GCTAAGCAAGCAAGCAACATCAGTCCCAATCAGTCTGTTGTCAGCAGTGGAACTGTG 45649  
 455 AAGTTGATGAGCAAGCAACATCAGTCTGTTGTCAGCAGTGGAACTGTG 514  
 45650 AAGTTGATGAGCAAGCAACATCAGTCTGTTGTCAGCAGTGGAACTGTG 45709  
 515 ACTCCCAACATAGTGTGTGGAAGATGCAAGTGTGCTGCTGTTTCAAGATTGCT 574  
 45710 ACTCCCAACATAGTGTGTGGAAGATGCAAGTGTGCTGCTGTTTCAAGATTGCT 45769  
 575 AGTACTTAAGGAGGCAAGCAACATCAGTCTGTTGTCAGCAGTGGAACTGTG 634  
 45770 AGTACTTAAGGAGGCAAGCAACATCAGTCTGTTGTCAGCAGTGGAACTGTG 45827  
 635 TATGTGAATTAATTTGAT-----GTGAAGCACTAATTAATTTGCTGAAG 681  
 45828 AGAGAGATTAATTAATTTGCTGCAAGAGTCTCAAACTTAAGTGTGCTTAAG 45887  
 682 TAAATTTCTTAATTAATTTTCTGTCAGACACTTTTCTAGACGTCGCAAGTTTGA 741  
 45888 TAAATGTAATTAATTTTCTGTCAGACACTTTTCTAGACGTCGCAAGTTTGA 45947  
 742 GCATGCGCTCTAAGAGCTTAAACCTA--TTTTTTACATGCTTAT--ATACATTC 797  
 45948 TTAATGCGCT--TAAAGCTTTTAAAGATGATTTTAACTCCCTTATCCATCAATTC 46006  
 798 CTAATGACATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 857  
 46007 CTACGACACCTCTCAAGTAATTTTA--ACATGATCTAGTACTAGTATTAATTAAT 46064  
 858 AACCCAGCCTATTGCAAAAATAAATCTTTTAAATTAATTAATTAATTAATTAATTAAT 917  
 46065 AACCCAGCCTATTGCAAAAATAAATCTTTTAAATTAATTAATTAATTAATTAATTAAT 46121  
 918 ATATACACTCTGGGAG 977  
 46122 AGATGAGCTCTG-----AGATGAGCTCTGAGTAAAGTAAATTTTCAAGAGAG 46172  
 978 AGGTGCGCTTTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1037  
 46173 AC-ATGCTGTTTACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 46225  
 1038 AGTTCAGTAATAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1097  
 46226 TGTTC-----ACTAGTTTAAAGTAAATTAATTAATTAATTAATTAATTAATTAAT 46270



QY 1098 ATATA 1102  
 Db 46271 ATATA 46275  
 RESULT 12  
 BC002687  
 LOCUS  
 DEFINITION BC002687 939 bp mRNA linear PRI 12-JUL-2001  
 Homo sapiens, NTF2-related export protein 1, clone MGC:3469  
 IMAGE:3608037, mRNA, complete cds.  
 ACCESSION BC002687  
 VERSION BC002687.1 GI:12803702  
 KEYWORDS MGC.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 939)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (05-FEB-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: ATCC  
 CNNA Library Preparation: Rubin Laboratory  
 CNNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: National Institutes of Health Intramural  
 Sequencing Center (NISC),  
 Gaithersburg, Maryland;  
 Web site: [http://www.nisc.nih.gov/nisc\\_mgcenhrgrl.nih.gov](http://www.nisc.nih.gov/nisc_mgcenhrgrl.nih.gov)  
 Contact: Shevchenko, Y., Wetherby, R.D., Beckstrom-Sternberg, S.M.,  
 Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,  
 Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,  
 Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCluskey, J.C.,  
 McDowell, J., Pearson, R., Snyder, B., Stantilpop, S., Thomas, P.J.,  
 Tlonsson, E.E., Touchman, J.W., Tsaurgeon, C., Vogt, J.L., Walker, M.A.,  
 Zhang, L.-H. and Green, E.D.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 12 Row: m Column: 11  
 This clone was selected for full length sequencing because it  
 passed the following selection criteria: matched mRNA gi: 7019470.  
 Location/Qualifiers  
 1. 939  
 /organism="Homo sapiens"  
 /db\_xref="LocustID:29107"  
 /db\_xref="taxon:9606"  
 /clone="MGC:3469 IMAGE:3608037"  
 /tissue\_type="uterus, endometrium adenocarcinoma"  
 /clone\_lib="NIH-MGC-44"  
 /lab\_host="DH10B-R"  
 /note="Vector: pOT87"  
 252..674  
 /codon\_start=1  
 /product="NTF2-related export protein 1"  
 /protein\_id="AA02687.1"  
 /db\_xref="GI:12803703"  
 /translation="MASVDFKTYVDACRAEEFVNYVTMDKRRRLSLRYMGAT  
 LVNAGNAVSGESLSEFFEMLPSEFOISVDCOPVDENTPQTVLVVIGCSYVFE  
 GNRQDFNQNFITLAQSPSNTWIKISDCFRFQDMAS"  
 BASE COUNT 217 a 253 c 250 g 219 t  
 ORIGIN  
 Query Match 22.0%; Score 251.6; DB: 9; Length 939;  
 Best Local Similarity 72.4%; Pred. No. 1.2e-46;  
 Matches 326; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

QY 162 CGCTCTGATTTTAAACTTAGTAGCAGCATGTAGCTGCTGAGAGTTGTCA 221  
 Db 256 CATCTGTGATTTCAAGACCTATGTGATCAGGCTGCAAGCTGCTGAGAGTTGTCA 315  
 QY 222 ATATTACTATGAGCAATGATTAAGAGAGAGGAGGACCTATACCTGATGACA 281  
 Db 316 ATGCTACTACACACCATGATAGAGGGGGGCTTGTGCTCCGCGCTGATGAGCA 375  
 QY 282 AGCCACCTTAATATGATGATGAAATGCTGTTTCAGGCTGAGTCCCTTAATATTTT 341  
 Db 376 CAGCCACCCCTGGTCTGGAATGCAATGCTGTTTCAGAGCAAGATCTTGAGAGATTTT 435  
 QY 342 TTGACACATTTGCTTGTAGTACCTCAGGTCAATATGATGATGCAACAGTTTCATG 401  
 Db 436 TTGAATTTGCTGCTTCCAGGAGTCTTCAATCAGCTGTGATGCTGACCTGTTTATG 495  
 QY 402 AGCAACCACTAGTCCCAACTACAGTCTTGTGTGACCACTGCAACTGTAAGTTTG 461  
 Db 496 ATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 555  
 QY 462 ATGGAACCAACACACATTTTTCACCAAGATTTCTGCTGATGCTCATCTCCCA 521  
 Db 556 AGGGGAACAACAACAGGAGCTTCAACCAAGATTTCTGCTGATGCTCATCTCCCA 615  
 QY 522 ACAATCTGTGTGAGATGATGCAAGTATGCTTCCAGTATGCTGATGTT 581  
 Db 616 GCAACACAGTGTGAGATGATGCAAGTATGCTTCCAGTATGCTGATGTT 675  
 QY 582 AAAGGGCAAAAGTCCATTTCTCATTTGTC 611  
 Db 676 GGGGGTGGCAGAGGTCTTGTGCTTCAATTC 705  
 RESULT 13  
 BC003029  
 LOCUS  
 DEFINITION BC003029 963 bp mRNA linear PRI 12-JUL-2001  
 Homo sapiens, NTF2-related export protein 1, clone MGC:4329  
 IMAGE:2820775, mRNA, complete cds.  
 ACCESSION BC003029  
 VERSION BC003029.1 GI:12804338  
 KEYWORDS MGC.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 963)  
 AUTHORS Strausberg, R.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-FEB-2001) National Institutes of Health, Mammalian  
 Gene Collection (MGC), Cancer Genomics Office, National Cancer  
 Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
 USA  
 REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
 COMMENT Contact: MGC help desk  
 Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
 Tissue Procurement: DCTD/DTF  
 CNNA Library Preparation: Rubin Laboratory  
 CNNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
 DNA Sequencing by: Genome Sequence Centre,  
 BC Cancer Agency, Vancouver, BC, Canada  
 Info@cgsc.bc.ca  
 Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
 Susanna Chan, Readman Chiu, Chris Fjell, Erin Garland, Ran Guin,  
 Leticia Hsiao, Martin Kirzysinski, Reta Kutsch, Oliver Lee, Soo  
 Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
 Ness, Pawan Pandoh, Anna-Liisa Prabhu, Parvaneh Saeedi, Jacqueline  
 Schein, Duane Smalhus, Michael Smith, Lorraine Spence, Jeff Stott,  
 Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,  
 George Yang, Scott Zuyderduyn, Marco Zerra.  
 Clone distribution: MGC clone distribution information can be found  
 through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 9 Row: p Column: 1





LOCUS BC000759 972 bp mRNA linear PRI 12-JUL-2001  
DEFINITION Homo sapiens, NFE2-related export protein 1, clone MGC:2701  
IMAGE:2820775, mRNA, complete cds.  
ACCESSION BC000759  
VERSION BC000759.1 GI:12653930  
KEYWORDS MGC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE 1 (bases 1 to 972)  
AUTHORS Strausberg, R.  
TITLE Direct Submission  
JOURNAL Submitted (15-NOV-2000) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,  
USA  
REMARK NIH-MGC Project URL: <http://mgc.ncl.nih.gov>  
COMMENT Contact: MGC help desk.

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: June 24, 2003, 19:25:38 ; Search time 78.8745 Seconds

(without alignments)  
5987.487 Million cell updates/sec

Title: us-09-763-902b-15

Perfect score: 2013

Sequence: 1 taccctccggagagagatggg.....ttttaagtaattttgtttt 1146

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 671580 segs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame\_n2p.model -DEV=xlp  
-Q/cgn2.1/USPTO/spool/US09763902/rumat.24062003.130341.12453/app.query.fasta.1.2254  
-DB=SPRMBL.21 -OPT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPCT=0  
-LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdt  
-LIST=45 -DOCALLIN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIG=15  
-MODE=LOCAL -OUTWRT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09763902.qcgn.1.1.232.qcgnat.24062003.130341.12453 -MCP=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

SPRMBL.21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_prodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriapi:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match length	ID	Description
1	167	8.3	200	Q9HDY6
				Q9hd6 schizosacch

2	163.5	8.1	167	5	Q8T2G7
3	150.	7.5	127	5	Q8T2D5
4	127.5	6.3	125	3	Q96VN3
5	121.5	6.0	123	3	Q9P8H0
6	120	6.0	146	10	Q8R032
7	114	5.7	688	5	Q917J5
8	114	5.7	690	5	Q9NH72
9	114	5.7	690	5	Q9VFR4
10	107.5	5.3	522	10	Q9LPR0
11	105	5.2	122	10	Q9EFK4
12	102.5	5.1	130	5	Q9VJ85
13	96	4.8	946	10	Q9SEV5
14	95	4.7	1378	5	Q9SYFO
15	94	4.7	537	5	Q21351
16	94	4.7	646	16	Q8X0B8
17	94	4.7	1052	10	Q49676
18	90.5	4.5	428	10	Q9LWX6
19	90.5	4.5	2151	12	Q9QD48
20	87.5	4.3	460	10	Q9FME2
21	87	4.3	1225	5	Q8T237
22	86	4.3	555	2	Q8VNR1
23	86	4.3	1595	2	Q52373
24	84	4.2	822	11	Q54923
25	83.5	4.1	427	10	Q9LDI9
26	83.5	4.1	596	4	Q9NTR6
27	83.5	4.1	757	4	Q8TR69
28	83.5	4.1	865	5	Q9N9S1
29	83.5	4.1	2151	12	Q9NRT5
30	83	4.1	304	3	Q13628
31	83	4.1	644	16	Q9AC71
32	83	4.1	872	16	Q9JSB4
33	83	4.1	872	16	Q9Z7H8
34	82.5	4.1	1253	3	Q59704
35	82.5	4.1	1454	12	Q66928
36	82.5	4.1	2151	12	Q9E127
37	82	4.1	280	16	Q8XNA7
38	82	4.1	507	8	Q9TK07
39	82	4.1	749	4	Q13433
40	82	4.1	753	12	Q8VSA9
41	81.5	4.0	579	12	Q9E1Y3
42	81.5	4.0	613	10	Q9MA69
43	81.5	4.0	815	12	Q9QRO3
44	81.5	4.0	6620	4	Q96AN2
45	81	4.0	334	10	Q9M255

## ALIGNMENTS

RESULT 1  
Q9HDY6 PRELIMINARY; PRT; 200 AA.

ID Q9HDY6  
AC Q9HDY6  
DT 01-MAR-2001 (TREMUREL.16, Created)  
DT 01-MAR-2001 (TREMUREL.16, Last sequence update)  
DT 01-JUN-2002 (TREMUREL.21, Last annotation update)  
DE Putative nuclear export factor.  
GN SPAB1A10.03.  
OS Schizosaccharomyces pombe (Fission Yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
OC Schizosaccharomycetales; Schizosaccharomycetaceae;  
OC Schizosaccharomycetes.  
OX NCBI\_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972H-;  
RA O'Neil S., Harris D., Wood V., Rajandream M.A., Barrell B.G.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AL512562; CAC21476.1;  
DR InterPro: IPR002075; NTF2.  
DR Pfam: PF02136; NTF2; 1.  
SQ SEQUENCE 200 AA; 23029 MW; B05E47D62B27B7FF CRC64;

Alignment Scores:

Pred. No.: 2.09e-09 Length: 200  
 Score: 167.00 Matches: 33  
 Percent Similarity: 53.39% Conservative: 30  
 Best Local Similarity: 27.97% Mismatches: 47  
 Query Match: 8.30% Indels: 8  
 DB: 3 Gaps: 3

US-09-763-902b-15 (1-1146) x Q9HDY6 (1-200)

OY 185 GTAGATCAGGATGTAGAGCTGAGAGTGTGTCAATATTCTATGAGACAATGGAT 244  
 Db 1 MetcUnserSerValLysTyrAlaGlnGluPheValGlnArgTyrSerSerLeuAsp 20  
 OY 245 AAAAGAGACGGGACGACACAGGCTGTATCTGACAGACAGCCACTTAATATGAAATGA 304  
 Db 21 ThrAsnArgAsnGlyIleAlaGluPheTyrArgGluAsnSerIleLeuTyrAsnGly 40  
 OY 305 AATGCTGTTCAGGGCTGGATGCCCTAAATATTTTGTGACACATTCCTTCTAGTAG 364  
 Db 41 LysProMetGln---ValThrGluPheThrSerMetIleValAsnLeuProTyrSerLys 59  
 OY 365 TTCACGTCATATGATGATGTCACCAACCACTTCATGACAGACACTGCCCAACT 424  
 Db 60 ThrLysValGluAspPheAspSerGln-----GlnValMetGlyAsnAspMet 75  
 OY 425 ACAGTCTCTTGTGACACAGTGAAGTGTGATGGAACAACAACATTTCTTC-484  
 Db 76 AsnIleIleIleValIleValSerGlyThrIleArgPheAspGlyLysLysProIleValPhe 95  
 OY 485 AACCAAGACTTCCTGCTGACTGCTCAGTCCACTCCCAACAATCTGTGTGAG 538  
 Db 96 SerTyrArg-----SerAlaThrThrGlnProAspPheThrIlePheLys 110

## RESULT 2

O8T2G7 PRELIMINARY: PRT: 167 AA.

AC 08T2G7  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical 19.0 kDa protein.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX4;  
 RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,  
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,  
 RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.,  
 RT "Sequence and Analysis of Chromosome 2 of Dictyostelium".  
 RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AC115607; AAL2371.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 167 AA; 19039 MW; E16EA5FADA00BAA CRC64;

## Alignment Scores:

Pred. No.: 4.92e-09 Length: 167  
 Score: 163.50 Matches: 40  
 Percent Similarity: 44.96% Conservative: 18  
 Best Local Similarity: 31.01% Mismatches: 68  
 Query Match: 8.12% Indels: 3  
 DB: 5 Gaps: 3

US-09-763-902b-15 (1-1146) x Q8T2G7 (1-167)

OY 176 AAAACTATGATGATCAGGATGTAGAGCTGTGAGAGTTGTC---AATATTACTAT 232  
 Db 38 LysLysIleValGluThrIleThrProArgAlaGluValIleIleLysGluPheTyr 57  
 OY 233 GAGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 292  
 Db 58 ProLysTyrAspSerSerArgAlaAspLeuIleGlyLeuTyrLysAspHisSerValSer 77

OY 293 ATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 352  
 Db 78 IleTyrAsnGlyThrGluCysLysGlyProGluHisIleGlyLysLeuAlaGluIle 97  
 OY 353 CCTTCTGTAGTACGTCAGGCTCAATATGTAGATGTCACCAACAGTTCATGACAGCAACT 412  
 Db 98 ProAsnSerValHisValIleGluThrPheAspAlaGlnProValProSerAsp---Asp 116  
 OY 413 CAGGCCCAACTACAGTCTGTGTGTGACGAGTGAAGTGTGATGATGATGATGATGATGAT 472  
 Db 117 LysGluAsnProAsnIleIleIleThrAlaThrGlyValIleThrLysThrSer 136  
 OY 473 CAACATTTCTTCAACAGCAAGTCTGTGATGATGATGATGATGATGATGATGATGATGATG 532  
 Db 137 GlnIleGlnPheHisGlnThrPheLeuValLysAspProThr---AsnSerAsnLeu 155  
 OY 533 TGAAGATGCAAGTATGCTTCCGT 559  
 Db 156 PheTyrLeuSerTyrAspCysIleArg 164

## RESULT 3

O8T2D5 PRELIMINARY: PRT: 127 AA.

AC 08T2D5  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Putative nuclear transport factor similar to nuclear transport factor 2 (Ntf2) from.  
 OS Dictyostelium discoideum (Slime mold).  
 OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
 OX NCBI\_TaxID=44689;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=AX4;  
 RA Gloeckner G., Eichinger L., Szafranski K., Pachebat J., Dear P.,  
 RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,  
 RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.,  
 RT "Sequence and Analysis of Chromosome 2 of Dictyostelium".  
 RL Submitted (APR-2002) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AC115613; AAM10776.1;  
 KW Hypothetical protein.  
 SQ SEQUENCE 127 AA; 14317 MW; 2D45B1B20021B298 CRC64;

## Alignment Scores:

Pred. No.: 1.43e-07 Length: 127  
 Score: 150.00 Matches: 37  
 Percent Similarity: 46.83% Conservative: 22  
 Best Local Similarity: 29.37% Mismatches: 55  
 Query Match: 7.45% Indels: 12  
 DB: 5 Gaps: 4

US-09-763-902b-15 (1-1146) x Q8T2D5 (1-127)

OY 191 CAGCATGTAGAGCTGCTGAGAGTGTGTCAATATTCTATGAGACAATGATGAAGA 250  
 Db 7 GlnValValGlyValGlyLysGlnPheValGlnHisTyrTyrGlyIlePheAspSerAsn 26  
 OY 251 AGACGGGACCTACACAGGCTGTATCTGACAGGACCACTTAATATGATGATGATGATGATG 310  
 Db 27 ArgAlaGlyLeuThrGlnIleTyrGlnGlnGlnThrIleThrIleThrIleThrIleThrIle 46  
 OY 311 GTTTCAGGCTGATGCCCTAAATATTTTGTGACATTTGCCCT---TCTAGTAG 364  
 Db 47 LeuSerGlyAlaAspAlaIleValIleHisIleValGluLeuProPheGlnGlnThrAsn 66  
 OY 365 TTCACGTCATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 424  
 Db 67 ArgLysIleAsnSerIleAspCysGlnGlnThrTyr-----GlnPro 80  
 OY 425 ACAGTCTTGTGTGACAGTGAAGTGTGAAGTGTGAAGTGTGAAGTGTGAAGTGTGAAGTGTG 484  
 Db 81 GlyIleMetIleThrValThrGlyThrLeuIleIleAspGlyGluAlaLys----- 97

```

QY 465 AACCCG---AACTTCTGCTGACTGCTAGTGCACACCAACAATCTGTGTGGAGATT 541
Db 98 ASngtLnLeuLysPpheValGlnValPheasnLeuAlaSerAsnAsnGlySerPheLeuLeu 117
QY 542 GCAAGTGAATGCTTCCT 559
Db 118 IleasnAspPheArg 123

RESULT 4
Q96VNV3 PRELIMINARY; PRT; 125 AA.
AC Q96VNV3
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Nuclear transport factor 2.
GN NTF2.
OS Emeritella nidulans (Aspergillus nidulans).
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eurotiales; Trichocomaceae; Emeritella.
CX NCBI_TaxID=5072;
RN [1]
RP SEQUENCE FROM N.A.
RA Espeso E.A., Penalba M.A.;
RT "Nuclear transport factor 2 from Aspergillus nidulans.";
RL Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY038983; AAK71467.1;
DR InterPro: IPR002075; NTF2.
DR Pfam: PF02136; NTF2".1
SQ SEQUENCE 125 AA; 13871 MW; 8FC18BDFC597E2F2 CRC64;

```

Alignment Scores:		4.29e-05	Length:	125
Pred. No.:	Score:	127.50	Matches:	36
Percent Similarity:	47.11%	Conservative:	21	
Best Local Similarity:	29.75%	Mismatches:	53	
Query Match:	6.33%	Indels:	11	
DB:	3	Gaps:	4	

US-09-763-902B-15 (1-1146) x Q96VN3 (1-125)

QY	206	GCTGAGAGATTGTCAATATTACTATGAGACATGGATAAAGAAGCGGCCTAAC	265
		:	:
Db	8	AlaIngInIpHeValThrPheYrYrGlnThrPheAspGlyAsnArgAlaGlyLeuAla	27
		:	:
QY	266	AGCGTGTCTCTGGACAAGGCCACTTAATATGATGAAATGCTCTTTCAGGCGCTGAT	325
Db	28	ProLeuYrYrAspAspHisSerMetLeuThrPheGlnThrSerAlaIleGlnGlyValAla	47
QY	326	GCCCTAAATTAATTTTGTGACACATTTGCT-----TTAGTGAAGTCCAGGCAATATG	379
Db	48	GlyIleIleGlnLysLeuThrSerLeuProPheGlnLysValGlnHisGlnValSerThr	67
QY	380	TTAGATTCCCAACCAAGTTCTATGACCAAGCACTAGTCCCAACTACAGTTCTGTGTGG	439
Db	68	LeuAspAlaGlnProSerGlyGlnHisGly-----GlyIleuValLeu	82
QY	440	ACCACTGGAACCTGTAAGTTTATGTGAACAACAACATTTCTTCACACAGAATCTCCTG	499
Db	83	ValThrGlyAlaLeuValAspGluGlnLysAsn-----ProMetAsnYrYrThr	99
QY	500	CTGACGTCTCAGTCCACTCCCAACATACT---GTGTGAGAGATTGCAGTGATTCGCTC	556
Db	100	GlnThrPheGlnLeuMetProAspGlyAlaGlySerYrPheValLeuAsnSpValPhe	119
QY	557	CGT 559	
Db	120	Arg 120	

RESULT 5  
 Q9P8H0 PRELIMINARY: PRT 123 AA.

```

DT 00P8H0: 00P8H0-2000 (TREMBLrel.. 15, Created)
DT 01-OCT-2000 (TREMBLrel.. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel.. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel.. 21, Last annotation update)
DE Rpl1.
GN Rpl1.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetæ;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RP Cordero Otero R.R., Lepingle A., Gallardin C.;
RT "Rpl1 gene from Yarrowia lipolytica."
RL Submitted (MAY-2000) to the EMBL/Genbank/DBDJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RP Rojas Quijano R., Lepingle A., Gallardin C.;
RT Submitted (APR-2000) to the EMBL/Genbank/DBDJ databases.
RL HML; AF260231; AAF70316.1;
DR HML; AF260231; IOUN.
DR HSSP; P13662; IOUN.
DR InterPro; IPR002075; NTF2.
DR Pfam; PF02136; NTF2; 1.
SO SEQUENCE 123 AA; 13696 MW; 3EECB95CF0058251 CRC64;

Alignment Scores:
Pred. No.: 0.000196 Length: 123
Score: 121.50 Matches: 37
Percent Similarity: 44.788 Conservative: 23
Best Local Similarity: 27.618 Mismatches: 57
Query Match: 6.048 Indels: 17
DB: 3 Gaps: 4

```

Alignment Scores:	pred. No.:	0	000196	Length:	123
Score:	121.50			Matches:	37
Percent Similarity:	44.78%			Conservative:	23
Best Local Similarity:	27.61%			Mismatches:	57
Query Match:	6.04%			Indels:	17
DB:	3			Gaps:	4

US-09-763-902B-15 (1-1146) x Q9P8H0 (1-123)
QY 164 TCCTGGATTTTAAACTATGTAGATCAGCCATGTAGAGTGGCTGAGAGATTGTCAAT 223
Db 2 SerValAspIhrheAsnIhr-----LeuAlaIySgInPhcysGlu 14
QY 224 ATTACTATGAGCAGCAATGATTAAGAAGACGGCAGTAAACCAGGCTGTATGTGACAAAG 283
Db 15 PheTyrTylGlnIhrheAspIhrAspArgSerGlnIleuGlyAsnIleuTyrArgAspHis 34
QY 284 GCCACGCTAATATGAAATGGAATGCGTGTTCAGGGCTGGAGTCCGCTAAATAATTTT 343
Db 35 SerMetIleuThrPheIhrGlyIhrGlnIhSgInGlyAlaIleValaIleValaIleVala 54
QY 344 GACACATGGCTTTCTAGTAG-----TTCCAGTCAATATGTAGATTGCCAACAGTT 397
Db 55 ValGlyIleuProPheGlyIleValaArgHisIySgIleSerAspIleAspAlaGlnProIa 74
QY 398 CAGGACGAAGCAAGTCAGTCCCAACACTACAGTCTTTGTGGACCAAGCAAGTGTGAAG 457
Db 75 SerAlaGlnIleGlyAsp-----ValIleValaIleValaIhrGlyIleuIcys 90
QY 458 TTGTATGGAACAACAACACATTTCTTCAACAGAACTTCCGTGACTGCTCAGTCCACT 517
Db 91 ValAspGlySerAspIhrProIeuProTyrGlyIleValPheHisIeu-----Ile 106
QY 518 CCCAACATACCTGTGTGAAGATTGCCAAGTCAATGTCCTCCGT 559
Db 107 ProAspGlySerSerTyrTyrValaPheAsnAspIlePheArg 120

RESULT 6	Q8RU32	PRELIMINARY:	PRT:	146 AA.
AC Q8RU32:				
DT 01-JUN-2002 (Tremblrel. 21, Created)				
DT 01-JUN-2002 (Tremblrel. 21, Last sequence update)				
DE Putative nuclear transport factor 2.				
EN P0415A04.24.				
OS Oryza sativa (japonica cultivar-group).				

OC Eukaryota: Viridiplantae: Streptophyta: Embryophyta: Tracheophyta;  
OC Spermatophyta: Magnoliophyta: Liliopsida: Poales; Poaceae;  
OC Ehnartoideae; Oryzaeae; Oryza.  
OC NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CV. NIPPONBARE;  
RA Sasaki T., Matsumoto T., Yamamoto K.;  
RT "Oryza sativa (japonica cultivar-group) genomic DNA, chromosome 1, PAC  
RT clone:P0415A04.";  
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL: AP003345; BAB90110.1; -  
SQ SEQUENCE 146 AA; 15796 MW; 4966E9A02E1E9670 CRC64;

Alignment Scores:	
Pred. No.:	0.000295
Score:	120.00
Percent Similarity:	44.53%
Best Local Similarity:	30.47%
Query Match:	5.96%
DB:	10
Gaps:	12
Length:	14
Matches:	39
Conservative:	18
Mismatches:	59
Indels:	12

US-09-763-902B-15 (1-1146) x Q8RU32 (1-146)

[illegible]

RESULT 7		
ID	091735	PRELIMINARY; PRF; 688 AA.
AC	091735	
DT	01-MAR-2001 (TREMBLrel. 16, Created)	
DT	01-MAR-2001 (TREMBLrel. 16, last sequence update)	
DT	01-JUN-2002 (TREMBLrel. 21, last annotation update)	
DE	Rin protein.	
GN	RIN OR CG9412.	
OS	Drosophila melanogaster (Fruit fly).	
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;	
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;	
OC	Ephydroidea; Drosophilidae; Drosophila.	
OX	NCBI_TaxID=7227;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN-BERKELEY;	
RX	MEDLINE=20196006; PubMed=10731132;	
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,	
RA	Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,	
RA	George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,	

RA Sutton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.-H.C., Blazey R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milius G.L.G.,  
RA Abill J.F., Abayaratne A., An H.-J., Andrews-Frankoch C., Baldwin D.,  
RA Bailew R.M., Basu A., Bexendale J., Bayraktiroglu L., Beasley E.M.,  
RA Beeson K.Y., Benos P.V., Bertram B.P., Bhandari D., Bolshakov S.,  
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brothier P.,  
RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Flitschmann W.,  
RA Foster C., Garfield A.E., Gaig N.S., Gebart W.M., Glasser K.,  
RA Glodex A., Gong F., Gotrell J.H., Gu Z., Guan P., Harris M.,  
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibeagwu C.,  
RA Jaitai M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Liao X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattle B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Merkulov G., Mishina N.V., Mobarry C., Morris J., Mosnir D.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Klimos I., Simpson M., Skupski M.P., Smith T.,  
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wassarman D.A., Weinstein G.M., Weissendach J.,  
RA Williams S.M., Woodgate T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveril J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of *Drosophila melanogaster*.";  
RL Science 287:2185-2195(2000).  
DR EMBL; AE003701; AAG22151.1; -;  
DR FlyBase; FBgn0015778; rin;  
DR InterPro; IPR002075; NTF2.  
DR InterPro; IPR000504; RNA\_rec\_mot.  
DR Pfam; PF02376; NTF2; 1.  
DR Pfam; PF00016; rrm; 1.  
DR SMART; SM00360; rrm; 1.  
DR PROSITE; PSS0102; RRM; 1.  
SQ SEQUENCE 688 AA; 74713 MW; 9E3299A99E7248C19 CRC64;

Alignment Scores:	
Pred. No.:	0.00178
Score:	114.00
Percent Similarity:	39.88
Best Local Similarity:	29.08
Query Match:	5.66
DB:	
	Length: 688
	Matches: 43
	Conservative: 16
	Mismatches: 63
	Indels: 26
	Gaps: 6

US-09-763-902B-15 (1-1146) x Q917J5 (1-688)

```

0Y      GCTAAAGTCCACAGATGGCCAGCGTCTCGAATTTTAAACATAATGAGATCAGGGCATG 199
      |||||.....:::
Db      AATnGlnSerGlnGlnProSerProGlnSer-----13
0Y      AGAGCTGCTGAGGAGGTTGTCTCAATATTTACTATGACACATGGATTAAGACAGCGGCA 255
      |||||.....|||
Db      14 ---ValGluArgGlnPheValArgGlnTyrTyrThrLeuLeuIleuSulSulAlaProAsnHis 32
      |||||.....|||
0Y      CTAACACAGGCTATCTGTGACACAGGCGCACCTTAATA---TGGAAATGGAATCGCTTTCA 316
      |||  |||  |||  |||  |||  |||
Db      33 LeuHisArgPheTyrAlaAsnHisAsnSerSerTyrIleHisGlyIleSerGlyLeuVal 52
      |||  |||  |||  |||  |||
0Y      GGGCGTGGATGGCCATAATATTTTGTGACACATGCGCTTCGATGATGTCACAGGCAAT 376
      |||  |||  |||  |||  |||
Db      53 GlyGlnArgGlnIleHisAsnArgIleGln-----GlnLeuAsn 65
      |||||
0Y      ATGTTAGATGTC-----CAACCGATTCATGACACAGCAACCTCAGTCCAAACT 424
      |||||  |||  |||||

```

```

Db      66  PheasnAerCySHnIsAlAlLysrIeSeerGInValAspAlaGlnAlaIhrIeUglYuanSgLy  85
QY      425  ACAcGTTCtTGTtGtGtGACCAAGTGGAAcTGTGAATtTGATGGAAACAACAACAcTTtTtTC  464
      |||||  |||||  |||  ::  |||||  |||
Db      86  ValValValGInValThr---GlyIuIueSerAsnAspGlyGInPrometaTargArG  104
QY      485  AACCAAGAAcTtCGtGtGACtGtGtCAAGtCCAcATCCAAACAATGtGTGGGAAGATtGCA  544
      |||  |||||:||||  |||||  |||  |||||  |||
Db      105  ThGtInthrPheValLLeuAlaIaGInSerProLyLys-----TyrTtYValHis  121
QY      545  AGtGATtGCTtCCGtTTTCAAGAT  568
      ::|||  |||||:|||||
Db      122  AsnAspIlePheArgTyrGInAsp  129

```

## RESULT 8

ID	Q9NH72	PRELIMINARY:	PRT:	690 AA.
AC	Q9NH72			
DT	01-OCT-2000 (TREMBLrel. 15, Created)			
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)			
DT	01-JUN-2002 (TREMBLrel. 21, Last annotation update)			
DE	Rasputin			
GN	RIN OR CG9412			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
CX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=20191744; PubMed=10725247;			
RA	Rasman C., Mayes C.A., Fanto M., Haynes S.R., Miodzik M.;			
RT	"Rasputin, the Drosophila homolog of the RasGAP SH3 binding protein,			
RT	functions in Ras and Rho mediated signaling."			
RL	Development 127:1715-1725(2000).			
DR	EMBL; AF231031; AAF68949.1; -			
DR	flyBase; FBgn0015778; rin.			
DR	InterPro; IPR002075; NTF2.			
DR	InterPro; IPR000504; RNA_rec_mot.			
DR	Pfam; PF02136; NTF2; 1.			
DR	Pfam; PF00076; rrm; 1.			
DR	SMART; SM00360; RRM; 1.			
DR	PROSITE; PS0102; RRM; 1.			
SO	SEQUENCE 690 AA; 74985 MW; 40C5BD7E95FA253 CRC64;			

### Alignment Scores:

Pred. No.:	0.00178	length:	690
Score:	114.00	Matches:	43
Percent Similarity:	39.86%	Conservative:	16
Best Local Similarity:	29.05%	Mismatches:	63
Query Match:	5.66%	Indels:	26
DB:	5	Gaps:	6

US-09-763-902B-15 (1-1146) x Q9NH72 (1-690)

140 GCTACAGGTCACAGATGCCACGTCCTGGATTTTAAACTATGTAGATCAGGCATGT 199

DD 9 AIAIINGINSELGINNINFLIOSELFIOGILINSEL 13

[illegible][illegible]

2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65  
66  
67  
68  
69  
70  
71  
72  
73  
74  
75  
76  
77  
78  
79  
80  
81  
82  
83  
84  
85  
86  
87  
88  
89  
90  
91  
92  
93  
94  
95  
96  
97  
98  
99  
100  
101  
102  
103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841

317

Db 55 G|vG|nArAG|uT|eH|sAsnArAT|eG|n-----G|nLeuAsn 67

377 ATGTTAGATTGC-----CAACCAGTTCATGAGCAAGCAACTCAGTCCCAACT 424

[illegible]

Db 68 Pheasnspscyshtsalalalyllseerclnvalnspalaglnalathrlheugclayasncl 87

QY 425 ACACTTCTTGTTGGACCACTGGACACTGTGAGTTGATGGAACAAACAACATTTCTC 484

Db 88 Valvalvalglnvalthr---glglulnueberasnspgylglnprometargargphe 106

QY 485 AACGAGAACTCCGCGACCTGCGACGTCGACGACCACTCCCAACATTACTGTGTGGAAGATGCA 544

Db 107 Thrslnthrphevalleualalalglhsrprolyslys-----tytytvalhis 123

QY 545 AGGATGTCCTCCGTTTCAAGAT 568

Db 124 Asnaspheargtyrclnasp 131

## RESULT 5

ID	Q9VFT4	PRELIMINARY;	PRT;	690 AA.
AC	Q9VFT4			
DT	01-MAY-2000	(TREMBLrel. 13, Created)		
DI	01-MAY-2000	(TREMBLrel. 13, Last sequence update)		
DI	01-JUN-2002	(TREMBLrel. 21, Last annotation update)		
DE	CG9412 protein (LID31194P).			
GN	RIN OR CG9412.			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_Taxid=7227;			
RN	[1]			

## RC STRAIN-BERKELEY;

RX MEDLINE-20196006. PubMed-10733132.  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Suton G.G., Mortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers J.H.C., Blaise R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abilaj J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Baller J.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Berson K.V., Benz P.V., Bernan B.P., Bhandari D., Bolshakov S.,  
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotler P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Doudna K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durkin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
 RA Fostel C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibejwam C.,  
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Kethum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Klup D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matzel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusslein D.R., Paefel J.M.,  
 RA Palazolo M., Pltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reihert K., Remington K., Saunders R.D.C., Scheele F., Shen H.,  
 RA Shue B.C., Slieden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Splier E., Spredling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svitskys R., Tecor C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissensbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaerit J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*."  
 SC The genome 287:2185-2195(2000).  
 [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;



RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaverl A.,  
RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,



RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gids R.A., Myers E.W., Rubin G.M., Venter J.C.,  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003659; AAF53669.1;  
 DR HSSP: P13662; I00N  
 DR FlyBase: FBgn0032680; CG10174.  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF02136; NTF2; 1.  
 SQ SEQUENCE 130 AA; 14582 MW; 39286BFC5BD7C34E CRC64;

## Alignment Scores:

Pred. No.: 0.0244 Length: 130  
 Score: 102.50 Matches: 35  
 Percent Similarity: 40.91% Conservative: 19  
 Best Local Similarity: 26.52% Mismatches: 57  
 Query Match: 5.09% Indels: 21  
 Gaps: 5

US-09-763-902b-15 (1-1146) x Q95FV5 (1-130)

QY 209 GAGAGCTTGTGATATTTACTAGACACATGAT-----AAAGAGAGCGGCA 259  
 Db :|||||  
 12 LysLupheValGlnGlnTYTYTAlaIlePheAspSpproAlaAsnArgLValVal 31  
 QY 260 CTACACAGCGTGTATGACAGCGCCACCTATATATGATGATGATGCTGTTCCAGG 319  
 Db :|||  
 32 IleAsnPhetYrAsnAlaThrAspSerPheMetThrPheGlnGlnGlnGlnGlnGln 51  
 QY 320 CTGAGTGCCTAATAATATTTTGGACACATGCTTCTAGTGCATCCAG-----370  
 Db :|||  
 52 AlaProLysIle-----LeuGlnLysValGlnSerLeuSerPheGlnLysIleAla 68  
 QY 371 -----GTCAATATGTTAGTATGTCACACACGTTTCATGACGACACACATCCCAACT 424  
 Db :|||  
 69 ArgValIleThrThrValAspSerGlnProThr-----SerAspGly 82  
 QY 425 ACAGTCTTGTGTGACACAGTGTGACAGTGTGATGATGATGATGATGATGATGATGAT 484  
 Db :|||||  
 83 GlyValLeuIleIleValLeuGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 102  
 QY 485 AACCAAGACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 544  
 Db :|||||  
 103 SerGlnIlePheLeu-----LysProAsnGlyLysLeuPheValAla 118  
 QY 545 AGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 580  
 Db :|||||  
 119 HisAspIlePheArgLeuAsnIleHisAsnSerAla 130

## RESULT 13

ID 095FV5 PRELIMINARY; PRT; 946 AA.  
 AC 095FV5;  
 DT 01-MAY-2000 (TREMBLrel. 13, Created)  
 DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Putative RNA-binding protein.  
 GN T1B9.8.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eucosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI-TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=CV. COLUMBIA;  
 RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,  
 RA Ronning C.M., Koo H., Fujii C.Y., Utterback T.R., Barnstead M.E.,  
 RA Bowman C.L., White O., Nierman W.C., Fraser C.M.,  
 RT "Arabidopsis thaliana chromosome III BAC T1B9 genomic sequence."  
 RT Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AC012395; AAF20221.1;  
 DR InterPro: IPR002075; NTF2.

DR InterPro: IPR000504; RNA\_rec.mot.  
 DR Pfam: PF02136; NTF2; 1.  
 DR Pfam: PF00076; rim; 2.  
 DR SMART: SM00360; RRM; 2.  
 DR PROSITE: PS0102; RRM; 2.  
 SQ SEQUENCE 946 AA; 105623 MW; AD4CD94A16D8AFD CRC64;

## Alignment Scores:

Pred. No.: 0.181 Length: 946  
 Score: 96.00 Matches: 35  
 Percent Similarity: 40.32% Conservative: 15  
 Best Local Similarity: 28.23% Mismatches: 60  
 Query Match: 4.77% Indels: 14  
 Gaps: 4

US-09-763-902b-15 (1-1146) x Q95FV5 (1-946)

QY 209 GAGAGCTTGTGATATTTACTAGACACATGATGATGATGATGATGATGATGATGATGAT 268  
 Db :|||||  
 283 AspLupheValArgGlnTYTYTAsnThrLeuGlnAsnAlaProGlnAsnLeuTyLys 302  
 QY 269 CTGATCTGACAGCGCCACCTTATATGATGATGATGATGATGATGATGATGATGATGATG 328  
 Db :|||||  
 303 LeuTyLysAspLysSerThrIle-----SerArgProGlyLeuAspGly 317  
 QY 329 CTAAATATTTT-----GACACATGCTCTGATGATGATGATGATGATGATGATGATGAT 376  
 Db :|||  
 318 ThrMetArgValPheThrLeuSerLysAspLeuLysTrpArgSerProGlySerPheAsp 337  
 QY 377 ATGTTAGATGCCAACACAGTTCATGACAGCAAGCAAGCAAGCAAGCAAGCAAGTCTGTT 436  
 Db :|||  
 338 SerValLysIleThrSerValThrSerGlnAspSerLeuLysGln-----GlyIleLeuVal 356  
 QY 437 GTGACAGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 496  
 Db :|||||  
 357 ValValTyLysIleTyLysLeuThrPheAsnGlnArgProAlaArgHisPheThrGlnValPhe 376  
 QY 497 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 556  
 Db :|||||  
 377 PheLeu-----ValProGlnGlnLysGlyTyLysIleValGlySerThrAspMetPhe 392  
 QY 557 CGTTTCAGAT 568  
 Db :|||||  
 393 ArgPheValAsp 396

## RESULT 14

ID 095FV5 PRELIMINARY; PRT; 1378 AA.  
 AC 095FV5;  
 DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE Hypothetical 153.1 kDa protein.  
 GN C07H6.3.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;  
 OC Rhabditidae; Felodertinae; Caenorhabditis.  
 OX NCBI-TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA MEDLINE=99069613; PubMed=9851916;  
 RA None;  
 RT "Genome sequence of the nematode *C. elegans*: a platform for  
 RT investigating biology. The *C. elegans* sequencing Consortium."  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=BRISTOL N2;  
 RA Macri C., Vaudin M.,  
 RT "The sequence of *C. elegans* cosmid C07H6."  
 RT Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.  
 RN [3]



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: June 24, 2003, 19:24:08 ; Search time 17.3988 Seconds  
(without alignments)  
5463.815 Million cell updates/sec

Title: US-09-763-902b-15  
Perfect score: 2013  
Sequence: 1-tacttcgggagagagatg99.....ttttaagtaatttttttt 1146

Scoring table: BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 112892 segs, 41476328 residues  
Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL-frame.n2p.spool -DEV-xlp  
-Q-/cgn2/\_USPTO.spool/US0976302/runtc\_24062003\_130340\_12418/app\_query.fasta.1.2254  
-DB-Swissprot.40 -QEMT-fastan -SUFFIX-n2p.rsp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS-bits -START=1 -END=1 -MATRIX-blosum62 -TRANS-human40.cdi  
-LIST=45 -DOCLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE-LOCAL -OUTMT=ptc -NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USR-US0976302.ecgn.1.1.46-ecgnal\_24062003\_130340\_12418 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -NAIT -DSPBLOCK=100 -LONGLOG  
-DEV-TIMEOUT=120 -WARN\_TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database: SwissProt.40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	749	37.2	142	1	NXT2_HUMAN
2	586	29.1	140	1	NXT1_MOUSE
3	584	29.0	140	1	NXT1_HUMAN
4	310	15.4	133	1	NXT1_DROME
5	207.5	10.3	137	1	NXT1_CAEEL
6	124	6.2	132	1	NFT2_ORYSA
7	115	5.7	127	1	NFT2_HUMAN
8	113	5.6	123	1	NFT2_SCHPO
9	112.5	5.6	126	1	NFT2_ARATH
10	108.5	5.4	124	1	NFT2_NEUCR
11	108	5.4	127	1	NFT2_XENTLA
12	104.5	5.2	124	1	NFT2_CANAL
13	101.5	5.0	125	1	NFT2_TENST
14	99.5	4.9	133	1	NFT2_CAEEL
15	88.5	4.4	482	1	G3B2_HUMAN
16	88.5	4.4	482	1	G3B2_MOUSE
17	87	4.3	643	1	NOS2_ALCEU
18	85	4.2	1452	1	VGL2_FITPV

C	19	81	4.0	358	1	VEGD_MOUSE	P97946 mus musculu
C	20	81	4.0	515	1	YNBT_YEAST	P53741 saccharomyc
C	21	81	4.0	579	1	UL25_YEVD	P09287 varicella-z
C	22	80.5	4.0	2151	1	RRPL_HANTV	P23456 hantaan vir
C	23	79	3.9	260	1	SVS1_YEAST	Q12254 saccharomyc
C	24	78.5	3.9	450	1	ACHX_HUMAN	Q99226 homo sapien
C	25	78.5	3.9	1176	1	SLAP_BACSH	P74290 bacillus sp
C	26	78	3.9	287	1	FFG_SYNY3	P41290 schizosacch
C	27	77.5	3.8	309	1	YDAA_HAEIN	P44195 haemophilus
C	28	77.5	3.8	530	1	UDBH_HUMAN	O75795 homo sapien
C	29	77.5	3.8	776	1	ARHG_HUMAN	Q15052 homo sapien
C	30	77.5	3.8	910	1	XO68_CAEEL	P34607 caenorhabdi
C	31	77.5	3.8	1451	1	VGL2_CVCAI	P36300 canine ente
C	32	77.5	3.8	2151	1	RRPL_SEOUB	P27314 seoul virus
C	33	77.5	3.8	778	1	PRTY_SCHPO	O42933 schizosacch
C	34	76.5	3.8	355	1	STBA_MOUSE	O64687 mus musculu
C	35	76.5	3.8	434	1	G3BP_SCHPO	O94260 schizosacch
C	36	76.5	3.8	4128	1	PRKD_HUMAN	O43915 homo sapien
C	37	76	3.8	740	1	CATE_STRRE	O87864 streptomyce
C	38	76	3.8	901	1	SOKI_YEAST	P40317 saccharomyc
C	39	75.5	3.7	626	1	RPOC_SYNY3	P74177 synechocyst
C	40	75.5	3.7	753	1	CKAA_BACUF	O32321 bacillus th
C	41	75	3.7	594	1	VEGD_HUMAN	O43915 homo sapien
C	42	75	3.7	594	1	KR2_HSVB	P28966 equine herp
C	43	74	3.7	375	1	YKUS_CAEEL	O17778 caenorhabdi
C	44	74	3.7	1112	1	RA13_SCHPO	P28706 schizosacch
C	45	73.5	3.6	363	1	VP43_NPVAC	P34050 autographa

## ALIGNMENTS

RESULT 1	NXT2_HUMAN	STANDARD	PRT	142 AA.
ID	NXT2_HUMAN			
AC	Q9NPJ8; Q9H8U0; Q9NRL7; Q9Y3M4; Q9Y3M5; Q9NQ64;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	NFT2-related export protein 2 (p15-2 protein) (DC9) (BM025).			
GN	NXT2.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OC	NCBI_TaxID=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).			
RC	TISSUE=Testis;			
RX	MEDLINE=20528640; PubMed=11073998;			
RA	Herold A., Suyama M., Rodrigues J.P., Braun I.C., Kuray U.,			
RA	Carmo-Fonseca C., Bork P., Izaurralde E.,			
RT	"Tap (NFT) belongs to a multigene family of putative RNA export			
RT	factors with a conserved modular architecture."			
RL	Mol. Cell. Biol. 23:8996-9008(2000).			
RN	[2]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RA	Kang Y., Cullen B.;			
RT	"p15-2, a homologous protein of p15, interacts with Tap."			
RL	Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.			
RN	[3]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RA	Li Y., Peng Y., Li N., Gu W., Han Z., Fu G., Chen Z.;			
RT	"Novel genes expressed in human dendritic cell."			
RL	Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A. (ISOFORM 1).			
RC	TISSUE=Ovarian carcinoma;			
RA	Isoqal T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,			
RA	Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,			
RA	Wagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,			
RA	Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,			
RA	Watanabe T., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,			
RA	Yamamoto T., Wakamatsu A., Nakamura Y., Nagahari K., Masuno Y.,			
RA	Ninomiya K., Iwayanagi T.;			







DR EMBL: BC003029; AA03029.1; -  
 DR EMBL: BC003410; AA03410.1; -  
 DR Genbank: HGNC:15913; NMT1.  
 DR MIM: 605811; -  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF02136; NTF2.1.  
 DR PROSITE: PS50177; NTF2\_DOMAIN.1.  
 KW Transport; Protein transport; Nuclear protein.  
 FT DOMAIN 16 NTF2.  
 FT SEQUENCE 140 AA; 15647 MW; 3586A86AC3944594 CRC64;  
 Alignment Scores:  
 Pred. No.: 1,51e-54 Length: 140  
 Score: 584.00 Matches: 104  
 Percent Similarity: 91.30% Conservative: 22  
 Best Local Similarity: 75.36% Mismatches: 12  
 Query Match: 29.01% Indels: 0  
 DB: 1 Gaps: 0  
 US-09-763-902b-15 (1-1146) x NMT1\_HUMAN (1-140)  
 QY 164 TCCTGATTTAACTTATGATGATCAGGATGAGCTGCTGAGAGTTGTCAT 223  
 Db 3 SerValAspPheLysTrpTrpValAspGlnAlaLacysArgAlaAlaGluGluPheValAsn 22  
 QY 224 ATTACTAGTACGATGATGATAAAGAAGAGCGGCGCTAACCGAGCTGATCTGAGACAG 283  
 Db 23 ValTrpTrpTrpTrpMetAspLysArgArgArgLeuLeuSerArgLeuTrpMetGlyThr 42  
 QY 284 GCGACCTTAATATGGAATGGAATGCTGTTTCAGGCGCTGATGCCCAATATATTTT 343  
 Db 43 AlaThrLeuValTrpAsnGlyAlaAlaValSerGlyGlnGlnSerLeuSerGluPhe 62  
 QY 344 GACACATTCCTTCTAGAGATTCAGGTCATATGTTAGTTGCCAACCGTTGATGAG 403  
 Db 63 GluMetLeuProSerSerGluPheGlnIleSerValAlaLacysGlnProValHisAsp 82  
 QY 404 CAAGCACTAGTCCCAACTACATCTCTGTTGACCACTGCACTGAGATTGAT 463  
 Db 83 GluAlaThrProSerGlnThrThrValLeuValValIleGlySerValLysPheGlu 102  
 QY 464 GGAACAACAACATCTTCTTACACCAAGACTCTGCTGAGTCTGACTGCTCAGTCCCAAC 523  
 Db 103 GlyAsnLysGlnArgAspPheAsnGlnAsnPhelLeuThrAlaGlnAlaSerProSer 122  
 QY 524 AATCTGCTGAGATGATGCAAGTGTCTCCGTTTCAAGTTGCTGAT 577  
 Db 123 AsnThrValTrpLysIleAlaSerAspCysPheArgPheGlnAspTrpAlaSer 140  
 RESULT 4  
 NMT1\_DROME  
 ID NMT1\_DROME STANDARD: PRT: 133 AA.  
 AC Q9V3H8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 GN NTF2-related export protein (p15).  
 GN NMT1 OR CG12752.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;  
 OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;  
 OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.  
 CX NCB1\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE-20036817; PubMed-10567585;  
 RA Black B.E., Levesque L., Holaska J.M., Wood T.C., Paschal B.M.;  
 RT "Identification of an NTF2-related factor that binds Ran-GTP and  
 RT regulates nuclear protein export.";  
 RL Mol. Cell. Biol. 19:8616-8624(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN-Berkeley;

RA MEDLINE-2016006; PubMed-10731132;  
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abril J.F., Abghyan A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolintinas S.,  
 RA Borokova D., Botchan M.R., Bouch J., Brockstein P., Brotler P.,  
 RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Durbin K.J., Evangelista C.C., Ferriz C., Fertler S., Fleischmann W.,  
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Hostin D., Houston K.A., Howland T.J., Hernandez J.R., Houck J.,  
 RA Jallat M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,  
 RA Laoko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacled J.M.,  
 RA Palazzolo M., Peltan G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrstkas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weisenbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yen R.-F., Zaveli J.S., Zhao M., Zhang G., Zhao Q., Zheng L.,  
 RA Zhang X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RL "The genome sequence of Drosophila melanogaster.";  
 Science 287:2185-2195(2000).  
 CC -1- FUNCTION: Stimulator of protein export for NES-containing  
 CC proteins. Also plays a role in the nuclear export of U1 snRNA,  
 CC tRNA, and mRNA (By similarity).  
 CC -1- SUBUNIT: Preferentially binds Ran-GTP (By similarity). Associates  
 CC with NMT1.  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: AF156959; AA054944.1; -  
 DR EMBL: AE003462; AA047066.1; -  
 DR FlyBase: FBgn0028411; NMT1.  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF02136; NTF2.1.  
 DR PROSITE: PS50177; NTF2\_DOMAIN.1.  
 KW Transport; Protein transport; mRNA transport; Nuclear protein.  
 FT DOMAIN 15 NTF2.  
 FT SEQUENCE 133 AA; 15181 MW; C6E644950AA370AA CRC64;  
 Alignment Scores:  
 Pred. No.: 4.15e-25 Length: 133  
 Score: 310.00 Matches: 57  
 Percent Similarity: 60.90% Conservative: 24  
 Best Local Similarity: 42.86% Mismatches: 48  
 Query Match: 15.40% Indels: 4  
 DB: 1 Gaps: 1





US-09-763-902B-15 (1-1146) x NTF2\_SCHPO (1-123)

QY 206 GCTGAGAGACTTTCATATTTACTATGAGACATGATATAAGACGCGCACTAAC 265  
||| :||| |||||:|||| ||| ||| :|||:  
Db 8 AlatrhrGlnPhehrnGlnPheTyrTyrGlnThrPheAspSerAspArgSerGlnLeuSer 27

```

OY 266 AGGCTGATCTGACAGAGCCACTTAATATGAAATGCTGTTACAGGCTGAT 325
    ||||| :||| :||| :||| :|||
DB 28 SerLeuTYRArgGluGlnSerMetLeuSerPheGlnGlnGlnGlnGlnGlnGln 47
OY 336 GCCCTAATATATTTTGTACACATTCCTCTAGTGAATCCAG-----GTCAATATG 379
    ||||| :||| :||| :||| :|||
DB 48 AlaIleValGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 67
OY 360 TTAGTTCCTCAACCACTTCATGACGACCACTGCTCCCAACTACATCTGTTGTG 439
    ||||| :||| :||| :||| :|||
DB 68 LeuSpAlaGlnPro-----ThrglyThrThrGlnGlnGlnGlnGlnGlnGln 82
OY 440 ACCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 496
    ||||| :||| :||| :||| :|||
DB 83 ValThrGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 102
OY 497 CTGCTGACTGCTGACATCCATCCCAATATCTGCTGGAAGATGAGTGAATGCTTC 556
    ||||| :||| :||| :||| :|||
DB 103 HisLeu-----ValAsnAsnAsnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 118
OY 557 CGT 559
    |||
DB 119 Arg 119

RESULT 9
NTEF2_ARATH STANDARD: PRT: 126 AA.
ID NTEF2_ARATH
AC 09C7F5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear transport factor 2 (NTEF-2).
GN NTEF2 OR A1162/970 OR F13K9.26.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OC NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv. Columbia;
RX MEDLINE=21016719; PubMed=11130712;
RA Theologis A., Ecker J.R., Palm C.J., Federspiel N.A., Kaul S.,
RA White O., Alonso J., Altieri H., Araujo R., Bowman C.L., Brooks S.Y.,
RA Buehler E., Chan A., Chen Q., Chen H., Cheuk R.F., Chin C.W.,
RA Chung M.K., Conn L., Conway A.B., Conway A.R., Creasy T.H., Dewar K.,
RA Dunn P., Eguu P., Feldblum T.V., Feng J.-D., Fong B., Fujii C.Y.,
RA Gill J.E., Goldsmith A.D., Haas B., Hansen N.F., Hughes B., Hultzer L.,
RA Hunter J.L., Jenkins J., Johnson-Hopson C., Khan S., Khaykin E.,
RA Kim C.J., Koo H.L., Kremenetskaia I., Kurtz D.B., Kwan A., Lam B.,
RA Langin-Hooper S., Lee A., Lee J.M., Lenz C.A., Li J.H., Li Y.-P.,
RA Lin X., Liu S.X., Liu Z.A., Luros J.S., Maiti R., Marziani A.,
RA Miltischer J., Miranda M., Nguyen M., Nierman W.C., Osborne B.I.,
RA Pai G., Peterson J., Pham P.K., Rizzo M., Rooney T., Rowley D.,
RA Sakano H., Salzberg S.L., Schwartz J.R., Shinn P., Southwick A.M.,
RA Sun H., Tallon L.J., Tambunga G., Toriumi M.J., Town C.D.,
RA Uteberck T., Van Aken S., Vaysberg M., Vysotskaia V.S., Walker M.,
RA Wu D., Yu G., Fraser C.M., Venter J.C., Davis R.W.
RA "Sequence and analysis of chromosome 1 of the plant Arabidopsis
RA thaliana".
RT Nature 408:816-820(2000).
RL
CC -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS. COULD BE
CC PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE
CC AT THE PORE COMPLEX DURING NUCLEAR IMPORT (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: CONTAINS 1 NTEF2 DOMAIN.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AC069471; AAC51491.1; -.
CC DR HSSP: P13662; LOUN.
CC DR InterPro: IPR002075; NTEF2.
CC DR Pfam: PF02136; NTEF2_1.
CC DR PROSITE: PS01077; NTEF2 DOMAIN: 1.
CC KM Transport; Protein transport.
CC FT DOMAIN 9
CC SO SEQUENCE 126 AA; 14002 MW; 263E06A7A8903E3B CRC64;

Alignment Scores:
Pred. No.: 0.000687 Length: 126
Score: 112.50 Matches: 33
Percent Similarity: 47.93% Conservative: 25
Best Local Similarity: 27.27% Mismatches: 52
Query Match: 5.59% Indels: 11
DB: 1 Gaps: 5

US-09-763-902b-15 (1-1146) x NTEF2_ARATH (1-126)
OY 206 GCTGAGAGTTTGTCAATATTTACTATGACCAATGATTAAGAAGAGGCACTAAC 265
    ||||| :||| :||| :||| :|||
DB 10 SerLysAlaPheValGlnHisIleTyrThrPheAspThrAsnArgValGlnGlnGln 29
OY 266 AGGCTGATCTGACAGAGCCACTTAATATGAAATGCTGTTACAGGCTGAT 325
    ||||| :||| :||| :||| :|||
DB 30 GlnLeuTYRArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 49
OY 336 GCCCTAATATATTTTGTACACATTCCTCTAGTGAATCCAG-----GTCAATATG 379
    ||||| :||| :||| :||| :|||
DB 50 SerIleValAlaLysLeuThrSerLeuProPheGlnGlnGlnGlnGlnGlnGlnGln 69
OY 380 TTAGTTCCTCAACCACTTCATGACGACCACTGCTCCCAACTACATCTGTTGTG 439
    ||||| :||| :||| :||| :|||
DB 70 ValAspGlnGlnPro-----SerGlyProAlaSerGlyMetLeuValPhe 84
OY 440 ACCAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 496
    ||||| :||| :||| :||| :|||
DB 85 ValSerGlnAsnLeuGlnLeuAlaGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 104
OY 497 CTGCTGACTGCTGACATCCATCCCAATATCTGCTGGAAGATGAGTGAATGCTTC 556
    ||||| :||| :||| :||| :|||
DB 105 HisLeu-----MetProThrProGlnGlnSerPheTyr--ValPheAsnAspIlePhe 121
OY 557 CGT 559
    |||
DB 122 Arg 122

RESULT 10
NTEF2_NEUCR STANDARD: PRT: 124 AA.
ID NTEF2_NEUCR
AC P87102;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear transport factor 2 (NTEF-2).
GN NTEF-2.
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariales; Sordariaceae; Neurospora.
OC NCBI_TaxID=5141;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BD-A / FGSC 1858;
RA Meyer U., Meyer M., Tschel D., Toeken K., Rensing L.;
RL Submitted (Mar-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS. COULD BE
CC PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE
CC AT THE PORE COMPLEX DURING NUCLEAR IMPORT (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: CONTAINS 1 NTEF2 DOMAIN.
CC

```

OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;





[illegible]

RC TISSUE-Brain;  
RA Kennedy D., Ru K., Matlick J.S.;  
RL Submitted (MAR-1998) to the EMBL/Genbank/DBJ databases.  
RM [5]  
RP SEQUENCE FROM N.A. (ISOFORM B).  
RC TISSUE-B-cell;  
RA Strausberg R.;  
RL Submitted (JUL-2001) to the EMBL/Genbank/DBJ databases.  
RM [6]  
RP INTERACTION WITH IKAPPABALPHA.  
RX MEDLINE-20549669; PubMed-10969074;  
RA Prigent M., Barlat I., Langen H., Dargemont C.;  
RT "Ikappabalpha and Ikappabalpha/NF-kappa B complexes are retained in  
RT the cytoplasm through interaction with a novel partner, RasGAP  
RT SH3-binding protein 2".  
RL J. Biol. Chem. 275:36441-36449(2000)  
CC -1- FUNCTION: PROBABLE SCAFFOLD PROTEIN THAT MAY BE INVOLVED IN MRNA  
CC TRANSPORT (POTENTIAL).  
CC -1- SUBUNIT: BINDS TO THE N-TERMINAL DOMAIN OF IKAPPABALPHA.  
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A (SHOWN HERE) AND B; ARE  
CC PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL: AF145284; AAD51937.1;  
DR EMBL: AB014560; BA31635.1;  
DR EMBL: AF051311; AAC15705.1;  
DR EMBL: AF053535; AAC95292.1;  
DR EMBL: BC011731; AAH11731.1;  
DR HSSP: P09651; 1HA1.  
DR InterPro: IPR002075; NTF2.  
DR InterPro: IPR000504; RNA\_rec\_mot.  
DR Pfam: PF00076; rrm; 1.  
DR Pfam: PF02136; NTF2; 1.  
DR SMART: SM00360; RRM; 1.  
DR PROSITE: PS50177; NTF2\_DOMAIN; 1.  
DR PROSITE: PS50102; RRM; 1.  
DR PROSITE: PS00030; RRM\_RNP\_1; FALSE\_NEG.  
DR Transport: RNA-binding; Alternative splicing.  
KW DOMAIN 11 133 NTF2.  
FT DOMAIN 331 409 RNA-BINDING (RRM).  
FT DOMAIN 134 223 GUU-RICH.  
FT DOMAIN 419 479 GLY-RICH.  
FT VARSPIC 243 275 MISSING (IN ISOFORM B).  
FT CONFLICT 267 267 S -> P (IN REF. 2).  
FT CONFLICT 359 359 E -> V (IN REF. 3).  
FT CONFLICT 460 460 M -> I (IN REF. 3).  
SQ SEQUENCE 482 AA; 5411 MW; 25456A3F1AAE218 CRC64;

## Alignment Scores:

Pred. No.: 0.342 Length: 482  
Score: 88.50 Matches: 34  
Percent Similarity: 42.40% Conservative: 19  
Best Local Similarity: 27.20% Mismatches: 63  
Query Match: 4.40% Indels: 9  
DB: 1 Gaps: 4

US-09-763-902B-15 (1-1146) x G3B2\_HUMAN (1-482)

OY 212 GAGTTGCAATATTACTAGACATGATAAAGAGACGGCACTACACGGCTG 271  
|||||  
DB 14 Gtupvevalnrglnttyrtyrthrleuenuenlysalaproglttyrleuhtsargpne 33  
272 TATCTGACAAAGGCCACTTATATGATGAAATGGAATGCTGTTTCAGCG----- 319

DB 34 TyrGIyArgaAnSerSerTyValhISGLyValaAsPaLaSerGIyLysProGInGu 53  
OY 320 ---CTGCATGCCCAATAATATTTTGTGACACATGCTCTAGTCAAGTCAAT 376  
DB 54 AlaValTYrGIyGlnAsnAspIleHISLysValLeuSerLeuAsnPheserGIuys 73  
OY 377 ATGTTAGATTGCCAACACAGTTTCATGAGCAGACACTGACCAACTACAGTCTTGT 436  
DB 74 HIsThrLysILeArgHISValaSPaLaHISaLThrLeuSerAsp---GIyValaVal 92  
OY 437 GTGACCACTGGAAGCTGTAAGTTGATGTGAACAAACAACATTTCTCAACCAGACTTC 496  
DB 93 GlnValMetGIyLeuLeuSerAsnSerGIyGlnProGluArgLysPhemetGlnThrPhe 112  
OY 497 CTGCTGACTGCTCAG---TCCACTGCCCAACATATCTGTGTGAAGATTGCCAAGTATGC 553  
DB 113 ValLeuAlaProGluGlySerValProAsnLys-----PheTyValhISAsnAspMet 130  
OY 554 TTCGCTTTCAGAT 568  
DB 131 PheaTgTYrGIuAsp 135

Search completed: June 24, 2003, 19:32:13  
Job time : 21.398 secs



OY 389 CAACCATTCATGAGCAAGCAACTACGCCAACAGTCTTGTGTGACCACTGGA 448  
||||| : : : : :  
Db 74 GlnProthrProasp-----SerCysIleIleSerMetValValGly 87  
OY 449 ACTGTGAAGTTTATGGAACAACAACATTTCTTCAACCAAGACTTCGTGACTGCT 508  
||||| : : : : :  
Db 88 GlnLeuYsaAlaAspGluAspProIleMetCylpheHisGlnMetPheLeu----- 105  
OY 509 CAGGCCACCTCCCAACATACGTGTGAGAGTTTGCAGTGTCTTCCCGT 559  
Db 106 -----LysAsnIleAsnAspAlaTrpValCysTTThrasnAspMetPheArg 120

## RESULT 2

T38039  
Probable nuclear transport factor 2 - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 15-Sep-2000  
C:Accession: T38039; T37728  
R:McDougal, R.C.; Rajandream, M.A.; Barrell, B.G.; Whitehead, S.; Churcher, C.M.  
submitted to the EMBL Data Library, August 1999  
A:Accession: T38039  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-123 <MCD>  
A:Cross-references: EMBL:AL109951; PIDN:CA853052.1; GSPDB:GN00066; SPDB:SPAC1B9.01c  
A:Experimental source: Strain 972h; cosmid C1B9  
R:Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Walsh, S.V.  
submitted to the EMBL Data Library, December 1995  
A:Reference number: Z21740  
A:Accession: T37728  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: SL\_77-123 <DEV>  
A:Cross-references: EMBL:Z68197; PIDN:CAA92380.1; GSPDB:GN00066; SPDB:SPAC1SF9.03c  
C:Genetics:  
A:Gene: SPDB:SPAC1B9.01c  
A:Map position: 1  
A:Introns: 3/1, 30/3; 54/3; 88/3  
C:Superfamily: yeast nuclear transport factor NTF2

## Alignment Scores:

Pred. No.:	0.00124	Length:	123
Score:	113.00	Matches:	34
Percent Similarity:	47.93%	Conservative:	24
Best Local Similarity:	28.10%	Mismatches:	51
Query Match:	5.61%	Indels:	12
DB:	2	Gaps:	4

US-09-763-902B-15 (1-1146) x T38039 (1-123)

OY 206 GCTGAGAGTTTGTCAATATTACTATGAGCAATGATAAAGAGCGGCACTAAC 265  
||||| : : : : :  
Db 8 AlaThrGlnPheThrGlnPheTyrGlnThrPheAspSerArgSerGlnLeuSer 27  
OY 266 AGGCTGTATCTGACAAAGCCACCTTAATATGAAATGCTGTTCAAGGCTGAT 325  
||||| : : : : :  
Db 28 SerLeuYrArgGlnGlnSerMetLeuSerPheGlnGlnLeuGlnGlyThrLys 47  
OY 326 GCCCTAATAATTTTTTTGACACATTCGCTTCTAGTGCAGTTCAG-----GTCAATATG 379  
||||| : : : : :  
Db 48 AlaIleValGlnLysLeuValSerLeuProPheGlnArgValGlnHisArgIleSerThr 67  
OY 380 TTAAGTCCCAACCAAGTTCATGAGCAAGCAACTAGTCCCAACTCAGTTCTGTTGTG 439  
||||| : : : : :  
Db 68 LeuAspAlaGlnPro-----ThrGlyThrThrGlySerValIleValMet 82  
OY 440 ACCAGTGAAGTGTGAGATTGATGAGAAACAACA-----CATTTCTTCAACGAGACTTC 496  
||||| : : : : :  
Db 83 ValThrGlyGlnLeuLeuAspGlnGlnMetAlaGlnArgTyrSerGlnValPhe 102  
OY 497 CTGCTGACTGCTCAGTCACATCCCAACAATACGTGTGGAAGATTCAGATGATTCCTC 556

Db 103 HisLeu-----ValAsnAsnAsnGlyAsnTyrTyrValLeuAsnAspLeuPhe 118  
OY 557 CGT 559  
|||  
Db 119 Arg 119

## RESULT 3

B86405

Probable nuclear transport factor 2 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse ear cress)  
C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001  
C:Accession: B86405  
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon  
Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,  
ansen, N.F.; Hughes, B.; Hulzar, L.  
Nature 408, 816-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maitl, R.; Marzia  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: B86405  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-126 <STO>  
A:Cross-references: GB:AE005172; NID:G11024877; PIDN:AA626961.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1  
C:Superfamily: yeast nuclear transport factor NTF2

## Alignment Scores:

Pred. No.:	0.00141	Length:	126
Score:	112.50	Matches:	33
Percent Similarity:	47.93%	Conservative:	25
Best Local Similarity:	27.27%	Mismatches:	52
Query Match:	5.59%	Indels:	11
DB:	2	Gaps:	5

US-09-763-902B-15 (1-1146) x B86405 (1-126)

OY 206 GCTGAGAGTTTGTCAATATTACTATGAGCAATGATAAAGAGCGGCACTAAC 265  
||||| : : : : :  
Db 10 SerLysAlaPheValGlnHisTyrTyrSerThrPheAspThrAsnArgValGlyLeuAla 29  
OY 266 AGGCTGTATCTGACAAAGCCACCTTAATATGAAATGCTGTTCAAGGCTGAT 325  
||||| : : : : :  
Db 30 GlyLeuYrGlnGlnAlaSerMetLeuThrPheGlnGlnGlnLysIleGlnGlyValGln 49  
OY 326 GCCCTAATAATTTTTTTGACACATTCGCT-----TTAGTGAAGTCCAGTCAATATG 379  
||||| : : : : :  
Db 50 SerIleValAlaLysLeuThrSerLeuProPheGlnGlnLysHisIleSerThr 69  
OY 380 TTAAGTCCCAACCAAGTTCATGAGCAAGCAACTAGTCCCAACTCAGTTCTGTTGTG 439  
||||| : : : : :  
Db 70 ValAspCysGlnPro-----SerGlyProAlaSerGlyMetLeuValPhe 84  
OY 440 ACCAGTGAAGTGTGAGATTGATGAGAAACAACAATTC-----TTCAACGAGACTTC 496  
||||| : : : : :  
Db 85 ValSerGlyAsnLeuGlnLeuAlaGlyLysIleHisAlaLeuLysPheSerGlnMetPhe 104  
OY 497 CTGCTGACTGCTCAGTCACATCCCAACAATACGTGTGGAAGATTCAGATGATTCCTC 556  
||||| : : : : :  
Db 105 HisLeu-----MetProThrProGlnGlySerPheTyr---ValPheAsnAspLeuPhe 121  
OY 557 CGT 559  
|||  
Db 122 Arg 122

## RESULT 4

H86248

[illegible]

```

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86111; MUID:21016719; PMID:1130712
A:Accession: H86398
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-122 <STO>
A:Cross-references: GB:AE005172; NID:9802547; PIDN:AAF9949.1; GSPDB:GN00141
A:Genetics:
A:Gene: F17L21.10
A:Map position: 1
C:Superfamily: yeast nuclear transport factor NTF2

Alignment Scores:
Pred. No.: 0.0091 Length: 122
Score: 105.00 Matches: 36
Percent Similarity: 45.90% Conservative: 20
Best Local Similarity: 29.51% Mismatches: 52
Query Match: 5.22% Indels: 14
DB: Gaps: 5

US-09-763-902B-15 (1-1146) x H86398 (1-122)
OY 206 GCTGAGGAGTTGTTCATATTTACTATAGAGCAATGATAAAGACGCGCAGCTAAC 265
    ||||| ||||| ||||| ||| ||| ||| |||
DB 7 AAlaySalapRevalIGluHISTYTYSerThrheaspalaasnaIproGlyLeuVal 26
    ||||| ||||| ||||| ||||| ||||| |||||
OY 266 AGGCTGATCTGCAACAGCCACCTATATAGAAATGCTCTTCAGGCGTGAT 325
    ||||| ||||| ||||| ||||| ||||| |||||
DB 27 SerLeuTYrGlnGluGlySerMetLeuThrPheGluGlnIlysIleGlnIleSerGln 46
    ||||| ||||| ||||| ||||| ||||| |||||
OY 326 GCCCTAAATATTTTGTGACATTTGCTCTTGTAGTCCAGGTCATATG----- 379
    ||||| ||||| ||||| ||||| ||||| |||||
DB 47 AsnIleValAlaIlyLeuThrGlyLeuProPheGlnIleGlnIlySHisAsnIleThrThr 66
    ||||| ||||| ||||| ||||| ||||| |||||
OY 380 TTGATATGCCACAGCTCATGACAGCAACAGCTAGGCCAACTAACAGTCTGTGTG 439
    ||||| ||||| ||||| ||||| ||||| |||||
DB 67 ValAspCysGlnPro-----SerIlyProAlaGlyIleMetLeuValPhe 81
    ||||| ||||| ||||| ||||| ||||| |||||
OY 440 ACCAGTGGAACTGAGATTGATGAACAACAACATTTTC-----TTCAACCAAC 493
    ||||| ||||| ||||| ||||| ||||| |||||
DB 82 ValSerGlyAsnLeuGlnLeuAlaGly----GluGlnHisAlaLeuIlyPheSerGlnMet 100
    ||||| ||||| ||||| ||||| ||||| |||||
OY 494 TTCCTGAGCACTGCTCAGTCCACATCCACACATATCTGTGAGAGATTGCAAGTGATGC 553
    ||||| ||||| ||||| ||||| ||||| |||||
DB 101 PheHisLeuIleSer-----AsnGlnIlyAsnTYrTYrValPheAsnAspIle 116
    ||||| ||||| ||||| ||||| ||||| |||||
OY 554 TTCGGT 559
    ||||| ||||| ||||| ||||| ||||| |||||
DB 117 PheArg 118

RESULT 6
S50467
Nuclear transport factor NTF2 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YER009w
C:Species: Saccharomyces cerevisiae
C:Date: 28-Jan-1995 #sequence_revision 12-May-1995 #text_change 24-Sep-1999
C:Accession: S50467; S41793; S72237
R:Dieterich, F.S.
submitted to the EMBL Data Library, December 1994
A:Description: The sequence of S. cerevisiae cosmid 9537, 9581, 9495, 9867, and lambd
A:Reference number: S50459
A:Accession: S50467
A:Molecule type: DNA
A:Residues: 1-125 <DIE>
A:Cross-references: EMBL:U18776; NID:9603592; PIDN:AAB64542.1; PID:9603601; MIPS:YER0
R:Haarer, B.K.; Petzold, A.S.; Brown, S.S.
submitted to the EMBL Data Library, July 1993
A:Description: Identification of mutations that are synthetically lethal with altered
A:Reference number: S41793
A:Accession: S41793
A:Molecule type: DNA
A:Residues: 1-84 <HAA>
A:Cross-references: EMBL:L22204; NID:9347714; PIDN:AAB49379.1; PID:9347715

```

R.Haerter, B.K.: Corbett, A.; Kweon, Y.; Petzold, A.S.; Silver, P.; Brown, S.S.  
Genetics 144, 495-510, 1996

A:Title: SEC3 mutations are synthetically lethal with profilin mutations and cause defec  
A:Reference number: S72237; MUID:970444444; PMID:8889515

A:Accession: S72237

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-84 <HMA>

A:Cross-references: EMBL:L22204; NID:g347714; PIDN:AA849379.1; PID:g347715

C:Genetics:

A:Gene: SGD:NPF2

A:Cross-references: SGD:S0000811; MIPS:YER009w

A:Map position: 5R

C:Superfamily: yeast nuclear transport factor NTF2

Alignment Scores:

Pred. No.:	0.0219	Length:	125
Score:	101.50	Matches:	38
Percent Similarity:	41.91%	Conservative:	19
Best Local Similarity:	27.94%	Mismatches:	58
Query Match:	5.04%	Indels:	21
DB:	2	Gaps:	6

US-09-763-902b-15 (1-1146) x S50467 (1-125)

```

OY 164 TCCTGATTTTAAACTTATGTAGATCAGCAGTACAGCTGAGAGTTTGTCAAT 223
    |||||
DB 2 SerLeuAspPheAsnThr-----LeuAlaGlnAsnPheThrGln 14
OY 224 ATTACTGTGAGACATGATTAAGAAGACGGCCTAACCGCTTATCTGGACAG 283
    |||||
DB 15 PheYrTYrAsnGlnPheAspThrAspArgSerGlnLeuGlnAsnLeuYrAsnGln 34
OY 284 GCCACCTTAATATGGAATGGAATGCTTTCAGGCGCGATGCCCTTAATATTTT 343
    |||||
DB 35 SerMetLeuThrPheGlnThrSerGlnLeuGlnGlnAlaLysAspIleValGlnLysLeu 54
OY 344 GACACATTCCTTACAGTGTCCAG-----GTCAATATGTGTGATGCCAACCACTT 397
    |||||
DB 55 ValSerLeuProPheGlnLysValGlnHisArgIleThrIleAspAlaGlnProAla 74
OY 398 CATGAGCAGCACTCACTCCCAACTACAGTCTTGTGTGACACAGTGAACCTGTGAG 457
    |||||
DB 75 SerProAsnGly-----AspValLeuValMetIleThrGlyAspLeuLeu 89
OY 458 TTGAT-----GGAAGAACACATTTCTTCAACCGCACTTCTCTGACTGCTCAG 511
    |||||
DB 90 IleAspGlnGlnGlnAsnProGlnArg---PheSerGlnValPheHisLeu----- 105
OY 512 TCACATCCCAACAATACATGCTGTGAGAGATTCGACAGTGTCTTCCTCGT 559
    |||||
DB 106 ---IleProAspGlyAsnSerTyrTyrValPheAsnAspIlePheArg 120

```

RESULT 7

hypothetical protein R05D11.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

C:Accession: T23921

R:White, S.

submitted to the EMBL Data Library, June 1996

A:Reference number: Z19818

A:Accession: T23921

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-133 <WII>

A:Cross-references: EMBL:Z75546; PIDN:CAA93890.1; GSPDB:GN00019; CESP:R05D11.3

A:Experimental source: clone R05D11

C:Genetics:

A:Gene: CESP:R05D11.3

A:Map position: 1

A:introns: 33/1; 63/3; 96/3

C:Superfamily: yeast nuclear transport factor NTF2

Alignment Scores:

Pred. No.:	0.0365	Length:	133
Score:	99.50	Matches:	34
Percent Similarity:	48.03%	Conservative:	27
Best Local Similarity:	26.77%	Mismatches:	47
Query Match:	4.94%	Indels:	19
DB:	2	Gaps:	6

US-09-763-902b-15 (1-1146) x T23921 (1-133)

```

OY 206 GCTGAGAGGTTTGCATATTACTATAGACATGAT-----AAAGA 250
    |||||
DB 11 AlaLysAlaPheIleGlnHisTyrTyrSerLysPheAspValGlyAspGlyMetSerArg 30
OY 251 AGACGGCCTAACACAGCCTGTAT---CTGACAAGCGCACCTTAATATGAATGAAT 307
    |||||
DB 31 AlaGlnGlyLeuSerAspLeuTyrAspProGlnAsnSerTyrMetThrPheGlnGlyGln 50
OY 308 GCTGTTTCAGGCTGATGCCCTAAATATTTTGTACACATTCCTTACAGTTC 367
    |||||
DB 51 GlnAlaLysGlyArgAspGlyIleLeuGlnLysPheThrIleGlnLysPheThrLysIle 70
OY 368 CAG-----GTCAATATGTGTGATGCCACACAGTTCATGAGCAACACACTCAGCCCA 421
    |||||
DB 71 GlnArgAlaIleThrValIleAspSerGlnProLeuTyrAspGlySerIleGln----- 88
OY 422 ACTACAGTTCCTTGTTCACACAGTGAACCTGATGATGGAACCAACACATTTTC 481
    |||||
DB 89 -----ValMetValLeuGlnLeuLysThrAspGlyAsnPro 104
OY 482 TTCACACGAACTTCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 538
    |||||
DB 105 PheSerGlnValPheIleLeu-----ArgProAsnAsnGlnLysSerTyrPhe 120
OY 539 ATTCGACGATGATGCTTCCT 559
    |||||
DB 121 IleGlyAsnGlnIlePheArg 127

```

RESULT 8

hypothetical protein K08F4.2 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999

C:Accession: T23479

R:Hendry, C.

submitted to the EMBL Data Library, January 1996

A:Reference number: Z19746

A:Accession: T23479

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-537 <WII>

A:Cross-references: EMBL:Z68879; PIDN:CAA93082.1; GSPDB:GN00022; CESP:K08F4.2

A:Experimental source: clone K08F4

C:Genetics:

A:Gene: CESP:K08F4.2

A:Map position: 4

A:introns: 66/3; 99/1; 140/3; 161/3; 338/1; 419/1

Alignment Scores:

Pred. No.:	0.185	Length:	537
Score:	94.00	Matches:	32
Percent Similarity:	46.88%	Conservative:	28
Best Local Similarity:	25.00%	Mismatches:	48
Query Match:	4.67%	Indels:	20
DB:	2	Gaps:	6

US-09-763-902b-15 (1-1146) x T23479 (1-537)

```

OY 203 GCTGCTGAGAG-----TTTGTCAATATTACTATGACACATGATTAAGA 250
    |||||
DB 96 AlaAlaGlnGlnValGlyGlyAlaPheCysHisGlnPheYrIleThrValSerGlnAsn 115

```





Db 19 phevalglulysfyrtyrhisnleuetyrlysserproserclnvalhsglnphetyr 38  
 QY 275 CTGGACAGAGCCACCTTAATATGGAATGGAAT-----CCTGTTTCAGGG 319  
 Db 39 leuaspaspservalleuclgylargproglyseraspclglumetvalservallysser 58  
 QY 320 CTGGATGCCCTAAATAATTTTGTGACACATTCCTTCAGAGTTCGAGTCGAATATG 379  
 Db 59 leuylsalailleasn-----gluglnilemetserpheasptryglulieserlys 75  
 QY 380 TTAGATTGCCAACCACTTCATGACGACGACATCCCAACTCACTTCCTGTCTG 439  
 Db 76 lileglnlleuethralaaspserglnalasertrymetasnclvalvalthrlleuval 95  
 QY 440 ACCAGTGAAGTCTGTAAGTTGTGGAACAAACAAACATTTCTTCAACGAGACTCTCG 499  
 Db 96 thnglyleuethrvallysglulgylargmetarg---pheserclnserphephe 114  
 QY 500 CTGACTGCTCAGTCCCTCCCAACATTAAGTGTGGAAGATTCGATGCTTCGCT 559  
 Db 115 leu-----valproleuasnnglysertryphevalleuasnaspvalphearg 130  
 QY 560 TTT 562  
 Db 131 tyr 131  
 RESULT 11  
 S24382  
 nitrous-oxide reductase (EC 1.7.99.6) - Alcaligenes eutrophus  
 C:Species: Alcaligenes eutrophus  
 C>Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 10-Sep-1999  
 C:Accession: S24382  
 R:Zumft, W.G.; Dreusch, A.; Loechehl, S.; Cuypers, H.; Friedrich, B.; Schneider, B.  
 Eur. J. Biochem. 208, 31-40, 1992  
 A:Title: Derived amino acid sequences of the nosz gene (respiratory N(2)O reductase) from  
 A:Implications for the Cu(A) site of N(2)O reductase and cytochrome-c oxidase.  
 A:Reference number: S24382; M0ID:92380183; PMID:1324835  
 A:Accession: S24382  
 A:Molecule type: DNA  
 A:Residues: 1-643 <ZUM>  
 A:Cross-references: EMBL:X65278; NID:938781; PIDN:CAA46383.1; PID:938782  
 C:Superfamily: nitrous-oxide reductase  
 C:Keywords: copper; metalloprotein; oxidoreductase  
 C:Keywords: copper; metalloprotein; oxidoreductase  
 F:624,624/Binding site: copper 1 (His, Cys, Cys) #status predicted  
 F:624,626,632/Binding site: copper 2 (Cys, His, Cys, His) #status predicted  
 Alignment Scores:  
 Pted. No.: 1,1 Length: 643  
 Score: 87.00 Matches: 40  
 Percent Similarity: 41.61% Conservative: 17  
 Best Local Similarity: 29.20% Mismatches: 62  
 Query Match: 4.32% Indels: 18  
 Db: 1 Gaps: 6  
 US-09-763-902B-15 (1-1146) x S24382 (1-643)  
 QY 137 ACTGTACAGCTCCAGATGGCAGCTCTGTGATTTT-----AAACTTAT 184  
 Db 361 ThrAlaThrValIleGlnleuSerArgValIleuGlyTyrPheaspclglysglnclulys 380  
 QY 185 GTAGATCAGCATGTAGACCTCTGAGAGTTT---GTCAATATTACTATGAGCAATG 241  
 Db 381 leuaspaspalaillevalnleuvalgluleuclgylproleuhsnlethralphe 400  
 QY 242 GATAAAGAGAGCGGACATAACGAGCTGTATCTGGACAGGCCACTTAATATGAT 301  
 Db 401 AspGlyArgGlyAspAlaIleTyrThrLeuPheLeuaspserclnleuvallystriphen 420  
 QY 302 GGAATATGCTTTTCAGGCTGATGCCGTAATATTTTTCAGACATTCCTCTACT 361  
 Db 421 -----leuaspalaillelelysphehislysglyasplysasnala 434  
 QY 362 GAGTTCAGGTCAATATATTAGATTGC-----CAACCAAGTTCATGACCAACCACTCAG 415

Db 435 lystyrvalvalaspargleuaspserleuclglntryglnproglyhisvalasnalaasercln 454  
 QY 416 TCCCAACACACAGTTCTGTGTGGACCACTGGACAGCTGAAGTTTATGTAACAAACAA 475  
 Db 455 SerGlnThr-----valalaalaaspclglylystyrleuvalalaglycyslys--- 470  
 QY 476 CATTTCTTCACAGCACTTCCTGCTGACTGCTGCATCCACATCCCAACAAT 526  
 Db 471 -----pheserlyaspargpheleuprovalglyproleuhsnproclunsn 486  
 RESULT 12  
 T31082  
 endo-1,4-beta-xylanase (EC 3.2.1.8) - Caldicellulositrupcor sp.  
 C:Species: Caldicellulositrupcor sp.  
 C>Date: 02-Sep-2000 #sequence\_revision 02-Sep-2000 #text\_change 02-Sep-2000  
 C:Accession: T31082  
 R:Morris, D.D.; Gibbs, M.D.; Ford, M.; Thomas, J.; Bergquist, P.L.  
 submitted to the EMBL Data Library, December 1997  
 A:Description: Family 10 and 11 xylanase genes from Caldicellulositrupcor sp. Rte9B.1.  
 A:Reference number: 220972  
 A:Accession: T31082  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1595 <MOR>  
 A:Cross-references: EMBL:AF036923; NID:92760904; PID:92760908; PIDN:AAB95325.1  
 C:Genetics:  
 A:Gene: xynB  
 C:Keywords: glycosidase; hydrolase  
 Alignment Scores:  
 Pted. No.: 1,66 Length: 1595  
 Score: 86.00 Matches: 32  
 Percent Similarity: 38.28% Conservative: 17  
 Best Local Similarity: 25.00% Mismatches: 50  
 Query Match: 4.27% Indels: 29  
 Db: 2 Gaps: 4  
 US-09-763-902B-15 (1-1146) x T31082 (1-1595)  
 QY 256 GGCACCTAACAGGCTGTATCTGACAGAGCCACCTTAATATGATGAAATGCTGTTTC 315  
 Db 730 Gylleuileaspclglyalaglyleuclnglyhisileasnvalasp-serproalavally 749  
 QY 316 AGGCGTGAATGCCCTAAATAATTTTGTGACACATTCGCTTCAGAGATTCAGATCA 375  
 Db 749 scluileglnuaspthrileasnleuapheserthrileproglyleuclnleuclnleth 769  
 QY 376 TATGTAGATTCGCAACACAGTTCATGACGACGACCACTCACTGATTCCTGT 435  
 Db 769 rgluleuasp-----ileserly 775  
 QY 436 TGTGACCACTGGAAGTGTGAGTTGATGGAACAAACAACATTTTC----- 481  
 Db 775 ltyrthrserserthrclnglnltyrtraspheleuproclnaspillemetileysglnal 795  
 QY 482 -----TTCACCAAGCACTTCCTGCTGACTGCTGCACAC-----CCCAACA 525  
 Db 795 aleuylsphelelysglnleuclnleuclnleuclnleuclnleuclnleuclnleucln 815  
 QY 526 TACTGTGTGAGATTCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 585  
 Db 815 lthrleuclnleuclnleuclnleuclnleuclnleuclnleuclnleuclnleuclnleucln 831  
 QY 586 GGGCAAAAGCTCATTCCTCATTT 607  
 Db 831 gserasntrproleuclnleuclnleuclnleuclnleuclnleuclnleuclnleuclnleuclnleucln 838  
 RESULT 13  
 VGIH79  
 E2 glycoprotein precursor - feline infectious peritonitis virus (strain 79-1146)  
 N:Alternate names: peplomer glycoprotein; spike glycoprotein  
 C:Species: feline infectious peritonitis virus





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_n2p model

Run on: June 24, 2003, 19:31:14 Search time 43.497 Seconds  
(without alignments)  
5701.778 Million cell updates/sec

Title: US-09-763-902b-15

Perfect score: 2013  
Sequence: 1 tacttcgagagagagatgg.....tttaagtaattttgtttt 1146

Scoring table:  
BLOSUM62  
Xgapop 10.0 Xgapext 0.5  
Xgapop 10.0 Xgapext 0.5  
Fgapop 6.0 Fgapext 7.0  
Delop 6.0 Delext 7.0

Searched: 417779 segs, 108206813 residues

Total number of hits satisfying chosen parameters: 835558

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08  
Maximum Match 1008  
Listing first 45 summaries

Command line parameters: -DEV-xlp  
-MODE=frame\_n2p.model -USO9-763902/runat\_24062003\_130343\_12522/app\_query.fasta\_1\_2254  
-DB=Published\_Applications\_AA -GFM=fastan -SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOPEL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR=SCORE=pcr -THR=MAX=100  
-THR=MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pcr -NOR=ext -HEARSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -USR=USO9763902.ecgn\_1\_1\_24\_etunat\_24062003\_130343\_12522  
-NCPU=6 -ICPU=3 -NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPELACK=100  
-LONGLOG -DEVTIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: Published Applications\_AA:

1: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PC7\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/PC7US\_PUBCOMB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	727	36.1	162	10	US-09-764-864-1180
2	584.5	29.0	164	10	US-09-764-864-1181
3	91.5	4.5	170	9	US-09-791-932-107
4	88.5	4.4	449	9	US-10-074-475-272

5	87	4.3	1454	10	US-09-854-799-22	Sequence 22, Appl
6	85	4.2	748 <td>10</td> <td>US-09-854-799-24</td> <td>Sequence 24, Appl</td>	10	US-09-854-799-24	Sequence 24, Appl
7	83.5	4.1	1127 <td>10</td> <td>US-09-815-242-5373</td> <td>Sequence 5373, Ap</td>	10	US-09-815-242-5373	Sequence 5373, Ap
8	83.5	4.1	1158 <td>10</td> <td>US-09-815-242-12522</td> <td>Sequence 12522, A</td>	10	US-09-815-242-12522	Sequence 12522, A
9	82	4.1	752	9	US-10-176-847-52	Sequence 52, Appl
10	81.5	4.0	756	9	US-10-077-130-5	Sequence 5, Appl1
11	81	4.0	358	9	US-09-852-209A-13	Sequence 13, Appl
12	81	4.0	358	9	US-10-131-600-13	Sequence 13, Appl
13	81	4.0	358	12	US-10-139-876-2	Sequence 13, Appl
14	79.5	3.9	1454	10	US-09-854-799-32	Sequence 32, Appl1
15	78.5	3.9	450	9	US-10-156-239-11	Sequence 11, Appl
16	78.5	3.9	450	9	US-10-199-485-11	Sequence 11, Appl
17	78.5	3.9	450	10	US-09-795-693-11	Sequence 11, Appl
18	78	3.9	748	10	US-09-854-799-28	Sequence 28, Appl
19	78	3.9	748	10	US-09-854-799-34	Sequence 34, Appl
20	78	3.9	1454	10	US-09-854-799-25	Sequence 26, Appl
21	76.5	3.8	3169	9	US-10-114-170-257	Sequence 257, Appl
22	76	3.8	1354	10	US-09-808-571A-4	Sequence 5, Appl1
23	75	3.7	351	9	US-09-829-378-5	Sequence 5, Appl1
24	75	3.7	354	9	US-09-375-248-6	Sequence 6, Appl1
25	75	3.7	354	10	US-09-956-095-2	Sequence 2, Appl1
26	75	3.7	354	10	US-09-219-345A-11	Sequence 11, Appl
27	75	3.7	354	10	US-09-795-006A-119	Sequence 119, Appl
28	74.5	3.7	488	9	US-09-951-217-42	Sequence 42, Appl
29	74	3.7	488	9	US-10-033-297-141	Sequence 141, Appl
30	74	3.7	488	9	US-09-940-244-141	Sequence 141, Appl
31	73.5	3.7	401	9	US-09-951-217-43	Sequence 43, Appl
32	73	3.6	362	12	US-10-139-876-4	Sequence 43, Appl1
33	73	3.6	576	10	US-09-925-297-787	Sequence 787, Appl
34	72.5	3.6	489	9	US-09-989-920-185	Sequence 185, Appl
35	71.5	3.6	390	9	US-09-966-782A-18	Sequence 18, Appl
36	71.5	3.6	1179	10	US-09-815-242-13608	Sequence 13608, A
37	71.5	3.6	1579	10	US-09-801-368-368	Sequence 368, Appl
38	71	3.5	266	9	US-09-738-626-6904	Sequence 6904, Ap
39	71	3.5	397	10	US-09-925-300-1531	Sequence 1531, Ap
40	71	3.5	457	9	US-09-877-963-2	Sequence 2, Appl1
41	70.5	3.5	253	9	US-09-880-748-1813	Sequence 1813, Ap
42	70.5	3.5	350	9	US-10-278-173-48	Sequence 48, Appl
43	70.5	3.5	619	10	US-09-801-368-88	Sequence 88, Appl
44	70.5	3.5	1668	9	US-10-223-070-15	Sequence 15, Appl
45	70	3.5	578	10	US-09-159-469-50	Sequence 50, Appl

## ALIGNMENTS

RESULT 1  
US-09-764-864-1180  
Sequence 1180, Application US/09764864  
Patent No. US20020132753A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT223  
CURRENT APPLICATION NUMBER: US/09/764,864  
CURRENT FILING DATE: 2001-01-17  
PRIOR application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1792  
SOFTWARE: Patentin Ver. 2.0  
SEQ ID NO 1180  
LENGTH: 162  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-764-864-1180

## Alignment Scores:

Pred. No.:	Score:	Percent Similarity:	Best Local Similarity:	Query Match:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
3	516-80	727.00	100.00%	36.12%	162	137	0	0	0	0

US-09-763-902b-15 (1-1146) x US-09-764-864-1180 (1-162)

QY	170	GATTTAAACTATATCAATCAGCAGCTACAGCTGCTGAGGAGTTGTCAATATTAC	229
Db	26	AspPheLysThrTyrValAspGlnAlaCysArgAlaIleArgLysPheValAsnIleTyr	45
QY	230	TATGAGACATGGATTAAAGAAGACGGGCACTTAACCAAGCTGTATCTGGACAAAGCCACC	289
Db	46	TyrGluThrMetAspLysArgArgAlaLeuThrArgLeuTyrIleAspLysAlaThr	65
QY	290	TTAATATGGAATGGAAATGCTGTTTCAGGGCTGATGCCCTTAATTAATTTTTTGACACA	349
Db	66	LeuIleTyrAsnGlyLysAlaValSerGlyLeuAspAlaLeuAsnAsnPheAspThr	85
QY	350	TTCGCTTCAGAGACTCCAGGCTCAATATGTAGATGCCCAACCAAGTCCATGACACACGA	409
Db	86	LeuProSerSerGlnPheGlnValAlaAsnMetLeuAspCysGlnProValHisGluGlnAla	105
QY	410	ACCACTATCCCAAACTACAGTCTTGTGTGACCACTGCAATGTCAGATTGATGGAAC	468
Db	106	ThrGlnSerGlnThrThrValLeuValValThrSerGlyThrValLysPheAspGlyAsn	125
QY	470	AAACAACATTCTTCACAGACAGACTTCCTGCTGACTGCATGCATCCCAACAATACT	529
Db	126	LysGlnHisPhePheAsnGlnAsnPheLeuThrAlaGlnSerThrProAsnAsnThr	143
QY	530	GTCGGGAAGTTCACAGTATGCTTCCTCCCTTTCAAAGTGTGCTACAGT	580
Db	146	ValTyrPylIleAlaSerAspCysPheAlaPheGlnAspTyrPheSerSer	162

```

RESULT 2
US-09-764-864-1181
: Sequence 1181, Application US/09764864
: Patent No. US20020132753A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PT223
: CURRENT APPLICATION NUMBER: US/09/764,864
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 1792
:
: SOFTWARE: PatentIn Ver. 2.0
:
: SEQ ID NO 1181

```

```

: TYPE: PRT
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: SITE
: LOCATION: (2)
: OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (3)
: OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (5)
: OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
: NAME/KEY: SITE
: LOCATION: (10)
: OTHER INFORMATION: xaa equals any of the naturally occurring L-amino acids
US-09-764-864-1181

```

Alignment Scores:	
Pred. No.:	1,15e-62
Score:	584.50
Percent Similarity:	90.85%
Best Local Similarity:	74.65%
Query Match:	29.04%
DB:	10
Length:	164
Matches:	106
Conservative:	23
Mismatches:	12
Indels:	1
Gaps:	1

US-09-763-902B-15 (1-1146) X US-09-764-864-1181 (1-164)

152 CAGATGGCCACGTCCTCTGGATTTAAACTTATGTAGATCAGGCATGTAGAGCTGCTGAG 211

```

Db      24  GtUmetAla-...SerValAspPheUlyThrTyValAspGlnAlaIaCysArGAlaAlaGlu 42
QY      212  GAGTTTGTCAAATTTACTATGAGCAATGCGTAAGAAGACGGGCATACCAGGCTG 271
Db      43  GlUpHeValAsnAlaTyTrTyTrThrThrMetAspLysArGArGLeUleUserArGLeU 62
QY      272  TATCTGACAAAGCCACCTTAATATGAAATGGAATGCTGTTTACAGGCGTCGATGCCCTA 331
Db      63  TyMeGcLyThAlaIaThrLeuValTrPAsnGlyAsnAlaValSerGlyGlnGluSerLeu 82
QY      332  AATAATTTTTTGACACATGCGCTCTAGTAGTCCAGCGCATATATGTTAGATTGCCAA 391
Db      83  SerGlUpPheHeUleUmetLeuProSerSerGluPheGlnIleSerValAlaAspCysGln 1021
QY      392  CCAGTTTCATGAGCAAGCAACTCAGTCCCAAACTACAGTTCCTGTGGACACAGTGCAGT 451
Db      103  ProValHIAspGlnAlaThrProSerGlnTrThrValLeuValAlaIleCysGlySer 1221
QY      452  GTGCAATTTGGAGGAACAACAACAATTTCTTCAACACGAAGCTTCGCTCAGTGCAG 511
Db      123  ValLyPheHeUleUlyAsnLysGlnArGAspPheAsnGlnAsnPheIleUeThAlaGln 1421
QY      512  TCCAGTCCCAACAATATCTGATGGAGAAGATGGCAATGATGCTTCCTCCGTTTCAAGATTGG 571
Db      143  AlAserProSerAsnThrValTrPylsIleAlaIaSerAspCysPheArGpHeGlnAspTrp 1621
QY      572  TCTAGT 577
Db      163  AlAser 164

```

```

RESULT 3
US-09-791-932-107
: Sequence 107, Application US/09791932
: Publication No. US20030003451A1
:
: GENERAL INFORMATION:
: APPLICANT: Vogell, Gabriel
: APPLICANT: Parodi, Luis A.
: APPLICANT: Hiebsch, Ronald R.
: APPLICANT: Lind, Peter
: APPLICANT: Kaytes, Paul S.
: APPLICANT: Ruff, Valerie
: APPLICANT: Huff, Rita M.
: APPLICANT: Wood, Linda S.
: TITLE OF INVENTION: No. US20030003451A1 G Protein-Coupled Receptors Cross-Refer
: FILE REFERENCE: 00325_US1

```

1 CURRENT FILING DATE: 2001-02-23  
2  
3 PRIOR APPLICATION NUMBER: 60/184,305  
4 PRIOR FILING DATE: 2000-02-23  
5  
6 PRIOR APPLICATION NUMBER: 60/184,304  
7 PRIOR FILING DATE: 2000-02-23  
8  
9 PRIOR APPLICATION NUMBER: 60/184,303  
10 PRIOR FILING DATE: 2000-02-23  
11  
12 PRIOR APPLICATION NUMBER: 60/184,357  
13 PRIOR FILING DATE: 2000-02-23  
14  
15 PRIOR APPLICATION NUMBER: 60/184,247  
16 PRIOR FILING DATE: 2000-02-23  
17  
18 PRIOR APPLICATION NUMBER: 60/188,880  
19 PRIOR FILING DATE: 2000-03-13  
20  
21 PRIOR APPLICATION NUMBER: 60/217,359  
22 PRIOR FILING DATE: 2000-07-11  
23  
24 PRIOR APPLICATION NUMBER: 60/217,370  
25 PRIOR FILING DATE: 2000-07-11  
26  
27 PRIOR APPLICATION NUMBER: 60/218,452  
28 PRIOR FILING DATE: 2000-07-20  
29  
30 PRIOR APPLICATION NUMBER: 60/186,810  
31 PRIOR FILING DATE: 2000-03-03  
32  
33 PRIOR APPLICATION NUMBER: 60/188,064  
34 PRIOR FILING DATE: 2000-03-09  
35  
36 PRIOR APPLICATION NUMBER: 60/186,457  
37 PRIOR FILING DATE: 2000-03-02  
38  
39 PRIOR APPLICATION NUMBER: 60/213,861  
40 PRIOR FILING DATE: 2000-06-23

```

P R I O R   F I L I N G   D A T E :   2001-02-13
N U M B E R   O F   S E Q   I D   N O S :   295
S O F T W A R E :   PatentIn version 3.1
S E Q   I D   N O   272
L E N G T H :   449
T Y P E :   P R T
O R G A N I S M :   Homo sapien
U S - 10-074-475-272

Alignment Scores:
Pred. No.:      0.161      Length:      449
Score:          88.50      Matches:      34
Percent Similarity: 42.408      Conservative: 19
Best Local Similarity: 27.208      Mismatches: 63
Query Match:    4.408      Indels:      9
DB:             9      Gaps:        4

US-09-763-902B-15 (1-1146) x US-10-074-475-272 (1-449)
OY 212 GAGTTTGCAATATTACTATGACACAAATGCAATAAAGACAGCGGCACTAACCGGCTG 271
14 GILPHPEVALARGILNITGTYTGRLEULEUAENLVALAPROGLIUTYLEUHNISARGPHE 33
OY 272 TACTGGAACAAGCCCACTTAATATAGGAATGGAATGCTGTTCAGG----- 319
34 TYGILYARGASERISERTYVALHISGLYGLVALAPALASERGLYLSPROGLINLU 53
OY 320 ---CTGAGATGCCCTAAATATTTTGTGACACATTCGCTCTAGTAGGATTCACAGTCAT 376
Db 54 ALAVALTYTGILYGLNINSMSPLEHNISLSTLYVALLEUSERLSEUANPHESEGLUCYS 73
OY 377 ATGTTAGATTGCCAACCACTTCATGAGCAAGCAACTGCCAAACTACAGTTCTGT 436
Db 74 HSTHLYLSILLEIGHISVALASPRALHNISLATHTHEUSERR--GLYVALVAL 92
OY 437 GTACCAAGTCGACACTGTGAGTTGATGGAACAACAACAATTTCTTCAACCAGACTTC 486
Db 93 GINVALMEGLYLEUUSERISERISERGLYGINPROGLIUALGYLSPHMETGLINTHPHE 112
OY 497 CTGCTAGACGCTGACG---TCCACGCTCCCAACAATACTGTGTGGAAGATTGCAAGTATGC 553
Db 113 VALLEUALPROGLIUSERVALPROASNLYS-----PHELYVALHNISASNPMET 130
OY 554 TTCGCTTTCAAGAT 568
Db 131 PHEARGTYRGLUASP 135

RESULT 5
US-09-854-799-22
Sequence 22, Application US/09854799
Patent No. US20020115064A1
GENERAL INFORMATION:
APPLICANT: SmithKline Beecham Corporation
TITLE OF INVENTION: Recombinant Feline Coronavirus S Proteins
NUMBER OF SEQUENCES: 54
CORRESPONDENCE ADDRESS:
ADDRESSEE: SmithKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King Of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-2799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/854,799
FILING DATE: 14-May-2001
CLASSIFICATION: <Unknown>
P R I O R   A P P L I C A T I O N   D A T A :

```

APPLICATION NUMBER: 08/392,459  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: SBC 14532B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5090  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ. ID NO: 22:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1454 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 22:  
US-09-854-799-22

Alignment Scores:  
Pred. No.: 0.402 Length: 1454  
Score: 87.00 Matches: 69  
Percent Similarity: 33.77% Conservative: 34  
Best Local Similarity: 22.62% Mismatches: 112  
Query Match: 4.32% Indels: 90  
DB: 10 Gaps: 17

US-09-763-902b-15 (1-1146) x US-09-854-799-22 (1-1454)

OY 260 CTACACGAGCTGATCTGACACAGCCACCTTAATGAAATGCTGTTTCAGG 319  
Db 211 ValThleuLeuYrserArGserThAlaThrTpoluYrserAlaAlaYrAla 230  
OY 320 CTGGATGCCCTAAATAATTTTTCACACATTCCTTCAGAGTTCAGTCAATAG 379  
Db 231 TyrlnglYAlserAsnPe-----ThrTyrlYrYlserAsnAsn 244  
OY 380 TTAGAT-----TGCCACAGTTCATGACGACAGCACTGCTGC 418  
Db 245 ThAsnGlyLeuYshThrYrGlulLeuCYsGluAsp---TyrluHlscYshThrYr 263  
OY 419 CAACACAGCTCTGTTGTGACGACGAGTGTGAGTGTGATGGAACAAACAACAT 478  
Db 264 AlAlThAsnAlhAlhAlaProThrserGlyGlyTrlLeProAspGly-----Phe 280  
OY 479 TTCTTCACACAGACTCTGCTGACTGCTGCTCAGT----- 517  
Db 281 SerPheAsnAsnTrpPheLeuLeuThrAsnSerSerThrPheValSerGlyArgPheVal 300  
OY 518 CCACAC-----AATACGTGTGGAAGTTGCAGT----- 547  
Db 301 ThAsnGlnProLeuLeuIleAsnCYsLeuTrpProAlProSerPheGlyAlaAla 320  
OY 548 -----GATGCTTCGCTTTTCAAGATGCTAGTAAAGGGCAAGTCCATTC 601  
Db 321 GlnGluPheCYsPheGluGlyAlaGlnPheSerGlnCYs-----AsnGlyAla 336  
OY 602 TCATTGGTCCATTAGTCCAGCAATGAATTAATTAATGATTAAT 646  
Db 337 SerLeuAsnAsnThrValAspValIleArgPheAsnLeuAsnPherThAlaAspValGln 356  
OY 647 -----TTT 649  
Db 357 SerGlyMetGlyAlaThrValPheSerLeuAsnThrThrGlyGlyAlaIleLeuGluIle 376  
OY 650 GATTGTAGACACATAATATATGCTGAACTAAATTTCTTAATATTTTATCTATC 709  
Db 377 SerCYsYrserAspThrAlaSerGluSerSer-----TyrSerYrGly 392  
OY 710 TCAGACACCTTTCTAGCAGCTGCCAGTGGAGCAATGCCCTCTA-----AGA 757  
Db 393 GluIleProPheGlyIleThrAspGlyProArgTyrCYsYrValLeuYrAsnGlyThr 412

OY 758 GCTTAAACATATTTTATACATGCTTAATATACATTCACATGACATCTTATATA 817  
Db 413 AlAlLeuYlYrleuGlyThrLeuPro-----ProSerValYlserIleAlaIle 429  
OY 818 AATTAACACATGATCTTGTGTAACATACATCTGACGCCATTCAGCAAAA 877  
Db 430 -----SerYlstrpGlyHlshPheYr-----IleAsnGly 439  
OY 878 TAAATCTTTTATATATATATATATGAGATGTCACACAAATATAC---ACCTGGGA 934  
Db 440 TyrAsnPheserThrPheProIleGlyCYs--IleSerPheAsnLeuThrThrGly 459  
OY 935 GAAGTGAGTTTGTGATTAATGTTAATTTCTAGTAAACACCTTGCTTTTCAG 994  
Db 459 lAserGlyAlaPheTrpThrIleAla---TyrThrserYrThrGluAlaLeuValGln 478  
OY 995 TT---AACACT 1002  
Db 478 AlGluAsnThr 481

RESULT 6  
US-09-854-799-24  
Sequence 24, Application US/09854799  
Patent No. US20020115064A1  
GENERAL INFORMATION:  
APPLICANT: SmithKline Beecham, Corporation  
TITLE OF INVENTION: Recombinant Feline Coronavirus S Proteins  
NUMBER OF SEQUENCES: 54  
CORRESPONDENCE ADDRESS:  
ADDRESS: SmithKline Beecham Corporation  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/854,799  
FILING DATE: 14-May-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/392,459  
FILING DATE: <Unknown>  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: King, William T.  
REGISTRATION NUMBER: 30,954  
REFERENCE/DOCKET NUMBER: SBC 14532B  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5090  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 748 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
SEQUENCE DESCRIPTION: SEQ ID NO: 24:  
US-09-854-799-24

Alignment Scores:  
Pred. No.: 0.537 Length: 748  
Score: 85.00 Matches: 69  
Percent Similarity: 33.77% Conservative: 34  
Best Local Similarity: 22.62% Mismatches: 112  
Query Match: 4.22% Indels: 90



DB: 10 Gap: 17

US-09-763-902B-15 (1-1146) x US-09-854-799-24 (1-748)

OY 260 CTAAACAGCCTGTATCTGTCAGACAGGCCACTTAATAATGAAGAATGCCTGTTCCAGGG 319  
:::||||| :||| ||| :|||  
Db 211 ValThrLeuLeuTyrSerArgSerSerThrAlaThrTrpGluTyrSerAlaAlaTyrAla 230  
:::||||| :||| ||| :|||  
OY 320 CTGATGCCCTTAATAAATTATTTTTTGACACATTCCTTCAGTAGAGTCCAGGCAATATG 379  
:::||||| :||| ||| :|||  
Db 231 TyrGlnGlyValSerAsnPhe-----ThyTyrTyrLysLeuAsn 244  
:::||||| :||| ||| :|||  
OY 380 TTAGAT-----TGCACAACCAGTTCATGACGACAGCAACTCACTCC 418  
:::||||| :||| ||| :|||  
Db 245 ThrAsnGlyLeuLysThrTyrGluLeucysGluAsp--TyrGluHisCysThrGlyTyr 263  
:::||||| :||| ||| :|||  
OY 419 CAAACTACAGTTCCTGTTGTGCACACAGTGAACGTGTAAAGTTATGGAACAACAACAT 478  
||| ||| ||| ||| ||| ||| |||  
Db 264 AlaThrAsnValPheAlaProThrSerGlyGlyTyrIleProAspIly-----Phe 280  
||| ||| ||| ||| ||| ||| |||  
OY 479 TTCCTCAACAGCACTCTCTGCTGACTGCTCACTCATCT----- 517  
||| ||| ||| ||| ||| ||| |||  
Db 281 SerPheAsnAsnTrpPheLeuLeuThrAsnSerSerThrPheValSerGlyArgPheVal 300  
||| ||| ||| ||| ||| ||| |||  
OY 518 CCCAAC-----AATACTGCTGGAAGATTCAGAT----- 547  
||| ||| ||| ||| ||| ||| |||  
Db 301 ThrAsnGlnProLeuLeuLeuLeuAsnCysLeuTrpProValProSerPheGlyValAlaAla 320  
||| ||| ||| ||| ||| ||| |||  
OY 548 -----GATTGCTCCGTTTCAAGATTGCTCAGTAGTTAAAGGGCAAAAGTCCATTG 601  
||| ||| ||| ||| ||| ||| |||  
Db 321 GlnGluPheCysPheGlnGlyValaGlnPheSerGlnCys-----AsnGlyVal 336  
||| ||| ||| ||| ||| ||| |||  
OY 602 TCATTGTGCTCCATTGTTGCCACCAATTGAATTTATGAAATTAT----- 646  
||| ||| ||| ||| ||| ||| |||  
Db 337 SerLeuAsnAsnThrValAlaSpValIleArgPheAsnLeuAsnPheThrAlaAlaValGln 356  
||| ||| ||| ||| ||| ||| |||  
OY 647 -----TTT 649  
||| ||| ||| ||| ||| ||| |||  
Db 357 SerGlyMetGlyAlaThrValPheSerLeuAsnThrThrGlyGlyValIleLeuGluIle 376  
||| ||| ||| ||| ||| ||| |||  
OY 650 GATTGTAGACGACATATATATGCTGTAACAAATTTCTTAATATTTCTATTCCTCG 709  
||| ||| ||| ||| ||| ||| |||  
Db 377 SerCysTrpSerAspThrValSerGluSerSerSer-----TyrSerTyrGly 392  
||| ||| ||| ||| ||| ||| |||  
OY 710 TCAGCACCTTTTCTAGACACCTGCCAGTTTGGAGCATTCGCCCTCA-----AGA 757  
||| ||| ||| ||| ||| ||| |||  
Db 393 GluIleProPheGlyIleThrAspGlyProArgTyrCysTyrValLeuTyrAsnGlyThr 412  
||| ||| ||| ||| ||| ||| |||  
OY 758 GCCTTAACAACTATTTTTTACATGCCCTTATATACATTCACATAGACATCTTATATA 817  
||| ||| ||| ||| ||| ||| |||  
Db 413 AlaLeuLysTyrLeuGlyThrLeuPro-----ProSerValLysGluIleAlaIle--- 429  
||| ||| ||| ||| ||| ||| |||  
OY 818 ATATTAACACATGATCTGTGACTACATACATGCACTGTAACCCAGCCGATTTGCCAAAA 877  
||| ||| ||| ||| ||| ||| |||  
Db 430 -----SerLysTrpGlyHisPheTyr-----IleAsnGly 439  
||| ||| ||| ||| ||| ||| |||  
OY 878 TAAAATCTTTTATATATATATATCTATGGATGACACACAATAATAC---ACTCTGGAA 934  
||| ||| ||| ||| ||| ||| |||  
Db 440 TyrAsnPhePheSerThrPheProIleGlyCys--IleSerPheAsnLeuThrThrGly 455  
||| ||| ||| ||| ||| ||| |||  
OY 935 GAAGTGGAGTTTITTTGCTATTAGGTTAAATTTTCTATGTAACACACGTTGCCGTTTTCAG 994  
||| ||| ||| ||| ||| ||| |||  
Db 459 AlSerGlyAlaPheThrPrpThrIleAla--TyrThrSerTyrThrGluAlaLeuValGlnV 478  
||| ||| ||| ||| ||| ||| |||  
OY 995 TT---AACAAT 1002  
||| ||| ||| ||| ||| ||| |||  
Db 478 AlGluAsnThr 481  
||| ||| ||| ||| ||| ||| |||

RESULT 7  
US-09-815-242-5373  
Sequence 5373, Application US/09815242  
Patent No. US20020061569A1  
GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert  
APPLICANT: Ohlson, Karl L.  
APPLICANT: Zyskind, Judith W.  
APPLICANT: Wall, Daniel  
APPLICANT: Trawick, John D.  
APPLICANT: Carr, Grant J.  
APPLICANT: Yamamoto, Robert T.  
APPLICANT: Xu, H. Howard  
TITLE OF INVENTION: Identification of Essential Genes in  
TITLE OF INVENTION: Prokaryotes  
FILE REFERENCE: ELTRA.011A  
CURRENT APPLICATION NUMBER: US/09/815,242  
CURRENT FILING DATE: 2001-03-21  
PRIORITY APPLICATION NUMBER: 60/191,078  
PRIORITY FILING DATE: 2000-03-21  
PRIORITY APPLICATION NUMBER: 60/206,848  
PRIORITY FILING DATE: 2000-05-23  
PRIORITY APPLICATION NUMBER: 60/207,727  
PRIORITY FILING DATE: 2000-05-26  
PRIORITY APPLICATION NUMBER: 60/242,578  
PRIORITY FILING DATE: 2000-10-23  
PRIORITY APPLICATION NUMBER: 60/253,625  
PRIORITY FILING DATE: 2000-11-27  
PRIORITY APPLICATION NUMBER: 60/257,931  
PRIORITY FILING DATE: 2000-12-22  
PRIORITY APPLICATION NUMBER: 60/269,308  
PRIORITY FILING DATE: 2001-02-16  
NUMBER OF SEQ ID NOS: 14110  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO: 5373  
LENGTH: 1127  
TYPE: PRT  
ORGANISM: Staphylococcus aureus  
US-09-815-242-5373

[illegible]





```

Db      193 gLycylsCysLeuProthrglyProArgHisProtySerIlelle-ArgArgSer11 212
QY      222 TTGACAAACTCCTACAGAGCTTACATGCGCTGATCTACATGTTTAAATCCAGAGAC 163
      |||
      |||
      |||
Db      212 e---GlnThrProGlnGlnsPrLcysProHisSerLysLysLeuGlyProIleAspme 231
QY      162 GTGGCCATCGGAGACCTTGAGACAGTGTCTTGCGGAGTCCGGCTCACTACT----- 108
      |||
      |||
      |||
Db      231 tLeuTrpAspAsnThrLysCysLysValLeuGlnAspGlnThrProLeuProGlyTh 251
QY      107 -----CCTCAC----- 102
      |||
      |||
      |||
Db      251 rGluAspHisSerTyrLeuGlnIncluprothLeuGlyProHisIleMetThrPheAsp61 271
QY      101 -----GTGTCCGGTCCGCA 88
      |||
      |||
      |||
Db      271 uAspArgCysGluCysValCysLysAlaProCysProGlyAspLeuIleGlnHisProG1 291
QY      87 GGGCTGGGAAATGATCAATCACTTATGAAAGAGAGCAAAAC---CGCCAAACGCAAAAT 31
      |||
      |||
      |||
Db      291 uAsnCysSerCysPheGluCysLysGlnSerLeuGlnSerCysGlnLysHisLys11 311
QY      30 TTTCACACCTCCCATTCCTCTCC 9
      |||
      |||
      |||
Db      311 ePheHisProAspThrCysSer 318

RESULT 12
US-10-131-600-13
; Sequence 13, Application US/10131600
; Publication No. US20030082670A1
GENERAL INFORMATION:
; APPLICANT: ERIKSSON, Ulf
; APPLICANT: ERIKSSON, Karin
; APPLICANT: LEE, Xuri
; APPLICANT: LEE, Xuri
; APPLICANT: PONTEN, Annika
; APPLICANT: UUTELA, Marko
; APPLICANT: ALITALO, Kari
; APPLICANT: OESTMAN, Arne
; APPLICANT: HELDIN, Carl-Henrik
; APPLICANT: BETSHOLTZ, Christel
; TITLE OF INVENTION: PLATELET-DERIVED GROWTH FACTOR C, DNA CODING
; TITLE OF INVENTION: THEREFOR, AND USES THEREOF
; FILE REFERENCE: 09-410349-Eriksson et al-1064-44740
; CURRENT APPLICATION NUMBER: US/10/131,600
; PRIOR APPLICATION NUMBER: US/09/410,349
; PRIOR FILING DATE: 1999-09-30
; PRIOR APPLICATION NUMBER: 60/108,109
; PRIOR FILING DATE: 1998-11-12
; PRIOR APPLICATION NUMBER: 60/110,749
; PRIOR FILING DATE: 1998-12-03
; PRIOR APPLICATION NUMBER: 60/113,002
; PRIOR FILING DATE: 1998-12-18
; PRIOR APPLICATION NUMBER: 60/135,426
; PRIOR FILING DATE: 1999-05-21
; PRIOR APPLICATION NUMBER: 60/144,022
; PRIOR FILING DATE: 1999-07-15
; NUMBER OF SEQ ID NOS: 39
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 13
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-131-600-13

Alignment Scores:
Pred. No.: 1.22 Length: 358
Score: 81.00 Matches: 38
Percent Similarity: 34.52% Conservative: 20
Best Local Similarity: 22.62% Mismatches: 67
Query Match: 4.00% Indels: 43
DB: 9 Gaps: 5

```

```

US-09-763-902B-15 (1-1146) x US-10-131-600-13 (1-358)
QY      387 CAATCTACATATATTGACCTGGAACTCAGTAGAAGCAATGTGTCAAAAATATTATTAG 328
      |||
      |||
      |||
Db      153 GlnGlnGlnValMetCysMetAsnThrSerThrSerTyrIleSerLysGlnLeuPheGlu 172
QY      327 GCATCCAGCCCTTAACACAGATTTCCATTCATTAATTAAGTGCCCTGTCCAGAAAGAC 268
      |||
      |||
      |||
Db      173 tIleSerValProLeuThrSerValProGlnLeuValProValLysIleAlaAsnHisThr 192
QY      267 -----CTGGTTAGTCCCGCTCTTTTATTCATGTGCTCATAGTAATA 223
      |||
      |||
      |||
Db      193 gLycylsCysLeuProthrglyProArgHisProtySerIlelle-ArgArgSer11 212
QY      222 TTGACAAACTCCTACAGAGCTTACATGCGCTGATCTACATGTTTAAATCCAGAGAC 163
      |||
      |||
      |||
Db      212 e---GlnThrProGlnGlnsPrLcysProHisSerLysLysLeuGlyProIleAspme 231
QY      162 GTGGCCATCGGAGACCTTGAGACAGTGTCTTGCGGAGTCCGGCTCACTACTACT----- 108
      |||
      |||
      |||
Db      231 tLeuTrpAspAsnThrLysCysLysValLeuGlnAspGlnThrProLeuProGlyTh 251
QY      107 -----CCTCAC----- 102
      |||
      |||
      |||
Db      251 rGluAspHisSerTyrLeuGlnIncluprothLeuGlyProHisIleMetThrPheAsp61 271
QY      101 -----GTGTCCGGTCCGCA 88
      |||
      |||
      |||
Db      271 uAspArgCysGluCysValCysLysAlaProCysProGlyAspLeuIleGlnHisProG1 291
QY      87 GGGCTGGGAAATGATCAATCACTTATGAAAGAGAGCAAAAC---CGCCAAACGCAAAAT 31
      |||
      |||
      |||
Db      291 uAsnCysSerCysPheGluCysLysGlnSerLeuGlnSerCysGlnLysHisLys11 311
QY      30 TTTCACACCTCCCATTCCTCTCC 9
      |||
      |||
      |||
Db      311 ePheHisProAspThrCysSer 318

RESULT 13
US-10-139-876-2
; Sequence 2, Application US/10139876
; Patent No. US20020123481A1
GENERAL INFORMATION:
; APPLICANT: Oliviero, Salvatore
; TITLE OF INVENTION: C-Fos Induced Growth Factor (Figf) And Dna Encoding Same
; FILE REFERENCE: 35784/205172
; CURRENT APPLICATION NUMBER: US/10/139,876
; CURRENT FILING DATE: 2002-05-07
; PRIOR APPLICATION NUMBER: 09/043,476
; PRIOR FILING DATE: 1998-03-18
; PRIOR APPLICATION NUMBER: PCT/IB96/0113
; PRIOR FILING DATE: 1996-09-30
; PRIOR APPLICATION NUMBER: GB9612368.2
; PRIOR FILING DATE: 1996-06-13
; PRIOR APPLICATION NUMBER: GB9519928.7
; PRIOR FILING DATE: 1995-09-29
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 358
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-139-876-2

Alignment Scores:
Pred. No.: 1.22 Length: 358
Score: 81.00 Matches: 38
Percent Similarity: 34.52% Conservative: 20
Best Local Similarity: 22.62% Mismatches: 67
Query Match: 4.00% Indels: 43
DB: 12 Gaps: 5

```

```

US-09-763-9028-15 (1-1146) x US-10-139-876-2 (1-358)
QY      387 CAATCTAACATATTGACCTCGGAAGCACTACTAGAGCCAAATGTGTCAAATAAATTATTAGG 328B
       : ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
Dg      153 GILGLUGLYALMETSrMetasnrThsrThrserrTyrrlleserLyslglnLeupheglu 172B
       : ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY      327 GCATTCAGGCCCTTAACACAGCATTTTCCATTCATTCATTAATTAAGGGCCCTTGCCAGATPACAGC 266B
       : ||| ||| ||||| ::::: ||| ::::: ||| ::::: ::::: :::::
Db      173 lIeserValProleuthrserValProgluleuValProvalLyslleaIsashsthr 192B
       : ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
QY      267 -----CTGGTAGTGCCCGCTCTCTTTATCCATGTCTCATTAATAATA 223B
       : ::::: ::::: ||| ::::: ::::: ::::: ::::: :::::
Db      193 GLyCyLyScySleupProthrghlyProahrghsPProflyrserllelle -Argaygserrl 212B
QY      222 TTGCACAACTCTCAGCACAGCTCTACATGCGTGATCTACATAGAATTAAAATCCAGAGAC 163B
       : ||||| ::::: ::::: ||| ::::: ||| ::::: :::::
Db      212 e--GlnthrProgluglaspGlucySpProhisSerlysylscysProileaspm 233B
QY      162 GTGGCCATCTGGACCTTTAGCAATGTGTTCGGCAATGCGCGTCACACTCT----- 108B
       : ||| ::::: ::::: ::::: ::::: ::::: |||||
Db      231 tleutPaapasmThrLysCyLyScyValleuInaspgluThrProleuprogllyrh 251B
QY      107 -----CCTCAC----- 102B
       : ::::: ::::: ::::: ::::: :::::
Db      251 rgluaSPHisserTyrlleuIngluProthrLeuCysglyProhisMetThrPheaspGl 271B
QY      101 -----GTGTCGGGTCCGCA 88B
       : ::::: ::::: ::::: ::::: :::::
Db      271 uAParGcySGlucyValcyLyAlaLaPrOcysProgllyaspLeulleInlsProgl 291B
QY      87 GGGCTCGGGAATGATCAATACTTAAGAAGAGCCAAACC---CGCCAAACGCCAAAT 31B
       : ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
Db      291 uaNCysSerCySphlegLucyLySglSerleuInserCyscysGlnLysHsLysIl 311B
QY      30 TTTCCACCCCTCATCTCTCC 9B
       : ||||| |||
Db      311 epheHsPProaspThrCysSer 318B

RESULT 14
US-09-854-799-32
Sequence 32, Application US/09854799
Patent No. US20020115064A1
GENERAL INFORMATION:
Applicant: SmithKline Beecham, Corporation
Title Of Invention: Recombinant Feline Coronavirus S Proteins
Number Of Sequences: 54
Correspondence Address:
Addresser: SmithKline Beecham Corporation
Street: 709 Swedeland Road
City: King of Prussia
State: PA
Country: USA
Zip: 19406-2799
COMPUTER READABLE FORM:
Medium Type: Floppy disk
Computer: IBM PC compatible
Operating System: PC-DOS/MS-DOS
Software: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
Application Number: US/09/854,799
Filing Date: 14-May-2001
Classification: <Unknown>
Prior Application Data:
Application Number: 08/3932,459
Filing Date: <Unknown>
Application Number: US 07/613,066
Filing Date: 14-NOV-1990
ATTORNEY/AGENT INFORMATION:
Name: King, William T.
Registration Number: 30,954
Reference/Docket Number: SBC 14532B
TELECOMMUNICATION INFORMATION:

```

	TELEPHONE: (215) 270-5015
	TELEFAX: (215) 270-5090
	INFORMATION FOR SEQ ID NO: 32:
	SEQUENCE CHARACTERISTICS:
	LENGTH: 1454 amino acids
	TYPE: amino acid
	TOPOLOGY: linear
	MOLECULE TYPE: protein
	SEQUENCE DESCRIPTION: SEQ ID NO: 32:
	US-09-854-799-32
	Alignment Scores:
	Pred. No.:           3, 36           Length:           1454
	Score:             79, 50          Matches:         77
	Percent Similarity: 34, 15%      Conservative:   35
	Best Local Similarity: 23, 48%    Mismatches:    121
	Query Match:       3, 95%       Indels:         95
	DB:                10           Gaps:           20
US-09-763-902B-15 (1-1146) x US-09-854-799-32 (1-1454)	
OY	209 GAGGAGTTTGTCTCAATATTAC-----TATGACACAATGGATAAA 247
Dd	187 Aaspqluphevalthralfatrylleserglyargsertryanstrpsnileasnasn 206
OY	248 AGAAGACGGGACTAACCCAGCGTGATTCGGACAAGGCCACTTAATATGGAATGGAAT 307
Dd	207 Trrpheasnasnalthrleuleuyserrargserterhralatrtpglutryser 226
OY	308 GCCTGTTCAAGCGCTGGATGCCTCAATATATTT-----TTTGCAC 346
Dd	227 Alalatlratyraltyrlnglyalvalsarnaphethrtyrtyrlysleuasnthrasn 246
OY	347 ACATTGCTCTTAGTAGAGTTCACAGTCAATATGTTAGATGCCAACCGATTACAGCAA 406
Dd	247 Glyleuysrthyrgluphe-----Cysgluasp--Tyrglutyr 259
OY	407 GCACATCAGTCCCAACTACAGTCTCTTGTCGACAGCGAGGACGTGACACTTGATGGA 466
Dd	260 Cystthelgyltiralathrasnsvalphelialprothvalglgyltyrllapropoly 279
OY	467 AACAAACAACTTTCTCAACACGAACTTCTGCTGACTGCTCATGCTCAC-----517
Dd	280 -----Pheaserpheasnasntprhleuleuthrasnsersetrhphervalser 296
OY	518 -----CCCAAC-----AATAGTGTGTGGAGATTGCAAGT--547
Dd	297 GlyarphavalthrassnglnproleuleuValasnqsluetrrppovalproserphe 316
OY	548 -----GATTGCTTCCTTTTCAAGATTGCTCTAGTAGTAAAGGGC 588
Dd	317 GlyValAlAlaInaglnIupheycspieguglyalaginPheSerGlnCys-----333
OY	590 AAAGTCACATTCATTTGGTCCATAGTTCACAGCAATGGAATTTATGCAATTAAT--646
Dd	334 ---SerclvalserleuasnasnthrValaspvallleaqrpheasnlleuasnphehr 352
OY	647 -----TTTGATTGTAGAAGCACT-----664
Dd	353 AlaaspyalginserglymetglyalatrValPheSerLeuasnthrhlglylVal 372
OY	665 -----ATAATATAGT-----GCTGAACCAATTTCTTAATATTT 700
Dd	373 IleleuclnualsercystyranspnhrValsergluseterpe-----Tyr 389
OY	701 CTATTCCTGTGCACACCTTTTCTAGACAGCTGCACATTGGACATTTGAGCACTCTA----754
Dd	390 SertrycljylultlepropheglylerhnrAspglyProArgtrycystyrValleutyr 409
OY	755 -----AGACCTTAACATATTTTTTAAACAGCCTTATATACATTCACATTAAGACAT 808
Dd	410 AsncglyThrAlauleuysrtyrleucglythrleupro-----ProseVallysglulle 427



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: June 24, 2003, 19:27:58 ; Search time 14.789 Seconds

(Without alignments)  
4559.976 Million cell updates/sec

Title: US-09-763-902B-15

Perfect score: 2013

Sequence: 1 taccctccggagagagatggg.....tttaagtaattttgtttt 1146

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame+.n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO\_spool/US09763902/runat\_24062003.130342.12495/app\_query.fasta\_1.2254  
-DB=Issued Patents\_AA -QFMT=fastin -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pcr -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptio -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09763902 -ECGN=1.1.42 -runat\_24062003.130342.12495 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEDUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

Issued Patents\_AA:\*  
1: /cgn2.6/ptodata/1/1aa/5A.COMB.pcp:\*  
2: /cgn2.6/ptodata/1/1aa/5B.COMB.pcp:\*  
3: /cgn2.6/ptodata/1/1aa/6A.COMB.pcp:\*  
4: /cgn2.6/ptodata/1/1aa/6B.COMB.pcp:\*  
5: /cgn2.6/ptodata/1/1aa/6C.COMB.pcp:\*  
6: /cgn2.6/ptodata/1/1aa/6D.COMB.pcp:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	4.3	1454	4	US-08-392-459-22
2	87	4.3	1454	5	PCT-US91-08525-22
3	87	4.3	1454	5	PCT-US93-04384-2
4	87	4.3	1454	5	PCT-US93-04384-47
5	86	4.3	1454	3	US-08-591-685-9
6	86	4.3	1454	5	PCT-US93-04384-44
7	86	4.3	1454	5	PCT-US93-04384-45
8	85	4.2	748	4	US-08-392-459-24
9	85	4.2	748	5	PCT-US91-08525-24
10	85	4.2	748	5	PCT-US93-04384-4
11	81	4.0	358	4	US-08-915-795-8
12	80.5	4.0	1454	5	PCT-US93-04384-43

13	80.5	4.0	1454	5	PCT-US93-04384-46	Sequence 46, Appl
14	80	3.9	321	4	US-08-915-795-9	Sequence 9, Appl
15	79.5	3.9	1454	5	US-08-392-459-32	Sequence 32, Appl
16	79.5	3.9	1454	5	PCT-US91-08525-32	Sequence 32, Appl
17	79.5	3.9	1454	5	PCT-US93-04384-12	Sequence 12, Appl
18	79.5	3.9	1454	5	PCT-US93-04384-48	Sequence 48, Appl
19	79	3.9	1168	4	US-08-620-171A-9	Sequence 9, Appl
20	78	3.9	748	4	US-08-392-459-28	Sequence 28, Appl
21	78	3.9	748	4	US-08-392-459-34	Sequence 34, Appl
22	78	3.9	748	5	PCT-US91-08525-28	Sequence 28, Appl
23	78	3.9	748	5	PCT-US91-08525-34	Sequence 34, Appl
24	78	3.9	748	5	PCT-US93-04384-6	Sequence 6, Appl
25	78	3.9	1454	4	US-08-392-459-26	Sequence 26, Appl
26	78	3.9	1454	5	PCT-US91-08525-26	Sequence 26, Appl
27	78	3.9	1454	5	PCT-US93-04384-8	Sequence 8, Appl
28	78	3.9	1454	5	PCT-US93-04384-16	Sequence 16, Appl
29	78	3.8	2325	3	US-08-417-089-6	Sequence 6, Appl
30	78	3.8	2325	4	US-08-695-651-6	Sequence 6, Appl
31	78	3.8	2325	4	US-08-930-285-6	Sequence 6, Appl
32	78	3.8	2325	4	US-08-695-421-6	Sequence 6, Appl
33	78	3.8	2325	4	US-08-697-826A-10	Sequence 10, Appl
34	77.5	3.8	530	4	US-09-180-852-2	Sequence 2, Appl
35	77.5	3.8	1451	1	US-08-308-872B-4	Sequence 4, Appl
36	76.5	3.8	3169	4	US-09-453-702B-257	Sequence 257, App
37	76	3.8	238	4	US-08-858-207A-420	Sequence 420, App
38	75	3.7	325	4	US-08-915-795-3	Sequence 3, Appl
39	75	3.7	354	4	US-08-915-795-5	Sequence 5, Appl
40	75	3.7	590	2	US-08-756-317-12	Sequence 12, Appl
41	75	3.7	1335	4	US-09-134-001C-3716	Sequence 3716, Ap
42	74.5	3.7	401	3	US-08-492-459-28	Sequence 28, Appl
43	74.5	3.7	401	4	US-08-716-873-42	Sequence 42, Appl
44	74.5	3.7	401	4	US-09-368-431-42	Sequence 43, Appl
45	74.5	3.7	401	4	US-09-414-006-28	Sequence 28, Appl

#### ALIGNMENTS

RESULT 1  
US-08-392-459-22  
; Sequence 22, Application US/08392459  
; Patent No. 6280974  
; GENERAL INFORMATION:  
; APPLICANT: SmithKline Beecham, Corporation  
; TITLE OF INVENTION: Recombinant Feline Coronavirus S  
; NUMBER OF SEQUENCES: 54  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: SmithKline Beecham Corporation  
; STREET: 709 Swedeland Road  
; CITY: King of Prussia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19406-2799  
COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/392,459  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/698,927  
; FILING DATE: 13-MAY-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/613,066  
; FILING DATE: 14-NOV-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: King, William T.  
; REGISTRATION NUMBER: 30,954  
; REFERENCE/DOCKET NUMBER: SBC 14532B  
; TELECOMMUNICATION INFORMATION:





380 TAGAT-----TGCACACAGTTCATGACAGCAACTGATCC 418  
245 ThrAsnGlyLeuLysThrTyrGluLeuGlySgluAsp---TyrGluHisCysThrGlyTyr 263  
419 CAACACTACAGTCTCTGTTGTGACAGCTGGAAGTTGAGTGAACAACAACAT 478  
264 AlaThrAsnValPheAlaProThrSerGlyTyrIleProAspGly-----Phe 280  
479 TTCTTCACACAGACTCTCTGCTGACTGCTCAGTCCACT----- 517  
281 SerPheAsnAsnThrPheLeuLeuThrAsnSerSerThrPheValSerGlyArgPheVal 300  
518 CCCAAC-----AATACTGCTGGAAGATTCGACAT----- 547  
301 ThrAsnGlnProLeuLeuLeuAsnCysLeuTrpProValProSerPheGlyValAlaAla 320  
548 -----GATTGCTCCGCTTTTTCAGATGCTGCTAGTAAAGGGCAAAAGTCATTC 601  
321 GlnGluPheCysPheGluGluGlyAlaGlnPheSerGlnCys-----AsnGlyVal 336  
602 TCATTGGTGCATTAATCTTCACAGCAATGAATTATGTAATTAT----- 646  
337 SerLeuAsnAsnThrValAspValIleArgPheAsnLeuAsnPheThrAlaAspValGln 356  
647 -----TTT 649  
357 SerGlyMetGlyAlaThrValPheSerLeuAsnThrThiGlyGlyValIleLeuGluIle 376  
650 GATTGTAGACAGCACTTAATATGCTGGAACATAATTCTTAATTTCTATTCCTG 709  
377 SerCysTyrSerAspThrValSerGluSerSer-----TyrSerTyrGly 392  
710 TCAGCACCTTTCTACACAGCTGCCAGCTTGGACATTCCTCTA-----AGA 757  
393 GlnIleProPheGlyIleThrAspGlyProArgTyrCysTyrValLeuTyrAsnGlyThr 412  
758 GCTTAAACTATTTTTCATGCTATATATATATATATATATATATATATATATATAT 817  
413 AlaLeuLysTyrLeuGlyThrLeuPro-----ProSerValLysGluIleAlaIle--- 429  
818 ATATTAAACACATGATCTGCTGCTACTACATCTGCTGACCCAGCTATTGCAAAA 877  
430 -----SerLysTrpGlyHisPheTyr-----IleAsnGly 439  
878 TAAATCTTTTAT 934  
440 TyrAsnPhePheSerThrPheProIleGlyCys--IleSerPheAsnLeuThrThiGlyAla 459  
935 GAAGTGGAGTTTGTGTTATAGTTAATTTCTAGTAAACACAGCTGCTGTTTTCAG 994  
459 IAserGlyAlaPheTrpThrIleAla---TyrThrSerTyrThrGluAlaLeuValGlnV 478  
995 TT---AACACT 1002  
478 AlaGluAsnThr 481

RESULT 3  
PCT-US93-04384-2  
Sequence 2, Application PC/TUS9304384  
GENERAL INFORMATION:  
APPLICANT: Miller, Timothy J.  
APPLICANT: Kieffer, Sharon  
APPLICANT: Reed, Albert Paul  
TITLE OF INVENTION: Compositions and Methods for Vaccination  
TITLE OF INVENTION: Against Coronaviruses  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SmithKline Beecham Corporation - Corporate  
STREET: 709 Swedeland Road  
CITY: King of Prussia

STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/04384  
FILING DATE: 19930507  
CLASSIFICATION:  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/882,171  
FILING DATE: 08-MAY-1992  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/698,927  
FILING DATE: 13-MAY-1991  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Schreck, Patricia A.  
REGISTRATION NUMBER: 33,777  
REFERENCE/DOCKET NUMBER: SBC H85009-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1454 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
PCT-US93-04384-2

Alignment Scores:  
Pred. No.: 0.0582 Length: 1454  
Score: 87.00 Matches: 69  
Percent Similarity: 33.77% Conservative: 34  
Best Local Similarity: 22.62% Mismatches: 112  
Query Match: 4.32% Indels: 90  
DB: 5 Gaps: 17

US-09-763-902b-15 (1-1146) x PCT-US93-04384-2 (1-1454)

260 CTAACACAGCTGTATCTGACCAAGCCACTTAATATGATGAATGAATGCTGTTTCAGG 319  
211 ValThrLeuLeuTyrSerArgSerThrAlaThrTrpGluTyrSerAlaAlaTyrAla 230  
320 CTGATGCCCTAATAATAATTTTGTGACACATGCTGCTTCTAGTGATTCAGGTCAGTCAATATG 379  
231 TyrGlnGlyValSerAsnPhe-----ThrTyrTyrLysLeuAsnAsn 244  
380 TTAGAT-----TGCACACAGTTCATGACAGCAACTGATCC 418  
245 ThrAsnGlyLeuLysThrTyrGluLeuGlySgluAsp---TyrGluHisCysThrGlyTyr 263  
419 CAACACTACAGTCTCTGTTGTGACAGCTGGAAGTTGAGTGAACAACAACAT 478  
264 AlaThrAsnValPheAlaProThrSerGlyTyrIleProAspGly-----Phe 280  
479 TTCTTCACACAGACTCTCTGCTGACTGCTCAGTCCACT----- 517  
281 SerPheAsnAsnThrPheLeuLeuThrAsnSerSerThrPheValSerGlyArgPheVal 300  
518 CCCAAC-----AATACTGCTGGAAGATTCGACAT----- 547  
301 ThrAsnGlnProLeuLeuLeuAsnCysLeuTrpProValProSerPheGlyValAlaAla 320  
548 -----GATTGCTCCGCTTTTTCAGATGCTGCTAGTAAAGGGCAAAAGTCATTC 601  
321 GlnGluPheCysPheGluGluGlyAlaGlnPheSerGlnCys-----AsnGlyVal 336

```

OY 602 TCATTGGTCATTAGTTCACCAATGAAATTATGTAATTAT----- 646
Db 337 SerLeuAsnAsnThValAspValIleArgPheAsnLeuAsnPhetThAlaAspValGln 356
OY 647 -----TTT 649
Db 357 SerGlyMetGlyAlaThValPheSerLeuAsnThrThrglyGlyValIleLeuGluIle 376
OY 650 GATTGTAGACACATATATATATGCTGTGAACATAATTCTTATATTTCTATCTCG 709
Db 377 SerCysTyrSerAspThValSerIleUserSerSer-----TysertYrGly 392
OY 710 TCAGACCTTTTCTAGACAGCTGCCAGTTTGAGCATTGCCCTCTA-----AGA 757
Db 393 GluIleProPheGlyIleThraspIyProArgTyrGlyTyrValLeuTyrAsnGlyThr 412
OY 758 GCTTTAAACATATTTTATACATGCTTATATATATATATATATATATATAT 817
Db 413 AlaLeuLysTyrLeuGlyThrLeuPro-----ProserVallyGluIleAlaIle--- 429
OY 818 ATATTAAACATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 877
Db 430 -----SerLysTyrGlyHisPheTyr-----IleasnGly 439
OY 878 TAAATCTTTTATATATATATATATATATATATATATATATATATATATATAT 934
Db 440 TyrAsnPhenPheSerThrPheProIleGlyCys---IleSerPheAsnLeuThrThrglyA 459
OY 935 CAATGAGGATTTTGGTTATAGTTATATTTCTAGTAAACACAGCTGCTGTTTTCAG 994
Db 459 laSerGlyAlaPheThrPheThrIleAla---TyrThrSerTyrThrGluAlaValGlnV 478
OY 995 TT---AACACT 1002
Db 478 alGluAsnThr 481

RESULT 4
PCT-US93-04384-47
Sequence 47, Application PC/TUS9304384
GENERAL INFORMATION:
APPLICANT: Miller, Timothy J.
APPLICANT: Klepfer, Sharon
APPLICANT: Reed, Albert Paul
APPLICANT: Jones, Elaine V.
TITLE OF INVENTION: Compositions and Methods for Vaccination
TITLE OF INVENTION: Against Coronaviruses
NUMBER OF SEQUENCES: 48
CORRESPONDENCE ADDRESS:
ADDRESSER: SmithKline Beecham Corporation - Corporate
ADDRESSER: Patents
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-2799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04384
FILING DATE: 19930507
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/882,171
FILING DATE: 08-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/698,927
FILING DATE: 13-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/613,066

```

```

FILING DATE: 14-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Schreck, Patricia A.
REGISTRATION NUMBER: 33,777
REFERENCE/DOCKET NUMBER: SBC H85009-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5015
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 47:
SEQUENCE CHARACTERISTICS:
LENGTH: 1454 amino acids
TYPE: AMINO ACID
MOLECULE TYPE: protein
PCT-US93-04384-47

Alignment Scores:
Pred. No.: 0.0562 Length: 1454
Score: 87.00 Matches: 69
Percent Similarity: 33.77% Conservative: 34
Best Local Similarity: 22.62% Mismatches: 112
Query Match: 4.32% Indels: 90
DB: 5 Gaps: 17

US-09-763-902b-15 (1-1146) x PCT-US93-04384-47 (1-1454)
OY 260 CTACACAGGCTGATCTGACACAGCCACCTTAATATGGAATGGAATGCTTTCAGGC 319
Db 211 ValThrLeuLeuTyrSerArgSerSerThralaThrTyrGluTyrSerAlaAlaTyrAla 230
OY 320 CTGATGCCCTAAATATTTTGTGACACATTCCTCTAGAGAGCTCAGCTCATATNG 379
Db 231 TyrGlnGlyValSerAsnPh-----ThrTyrTyrLysLeuAsnAsn 244
OY 380 TTAGAT-----TGCCACAGCTCATGACACAGCACTGATGC 418
Db 245 ThrAsnGlyLeuLysThrTyrGluLeuGlyCysGluAsp---TyrGluHisCysThrGlyTyr 263
OY 419 CAATGACAGTCTCTGTGTGACCGACGAGACCTGCAAGTTGATGGAACAAACACAT 478
Db 264 AlaThrAsnValPheAlaProThrSerGlyTyrIleProAspGly-----Phe 280
OY 479 TTCTTCACACAGAACTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 517
Db 281 SerPheAsnAsnThrPheLeuLeuThrAsnSerSerThrPheValSerClyArgPheVal 300
OY 518 CCCAAC-----AATACTGTGTGGAAGATTGCAAGT----- 547
Db 301 ThrAsnGlnProLeuLeuIleAsnGlyLeuTyrProValProSerPheGlyValAlaAla 320
OY 548 -----GATGCTTCCTGCTTTCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 601
Db 321 GlnGluPheCysPheGluGlyAlaGlnPheSerGlnCys-----AsnGlyVal 336
OY 602 TCATTGGTCATTAGTTCACCAATGAAATTATGTAATTAT----- 646
Db 337 SerLeuAsnAsnThValAspValIleArgPheAsnLeuAsnPhetThAlaAspValGln 356
OY 647 -----TTT 649
Db 357 SerGlyMetGlyAlaThValPheSerLeuAsnThrThrglyGlyValIleLeuGluIle 376
OY 650 GATTGTAGACACATATATATGCTGTGAACATAATTCTTATATTTCTATCTCG 709
Db 377 SerCysTyrSerAspThValSerIleUserSerSer-----TysertYrGly 392
OY 710 TCAGACCTTTTCTAGACAGCTGCCAGTTTGAGCATTGCCCTCTA-----AGA 757
Db 393 GluIleProPheGlyIleThraspIyProArgTyrGlyTyrValLeuTyrAsnGlyThr 412
OY 758 GCTTTAAACATATTTTATACATGCTTATATATATATATATATATATATATAT 817
Db 413 AlaLeuLysTyrLeuGlyThrLeuPro-----ProserVallyGluIleAlaIle--- 429

```

```

QY      818  ATATTAACACATGATCTGGTACTAATCACTAGTGAACCCAGCCTATTGCAAAA 877
Db      430  -----SerLysTrpIleAlaHisPheTyr-----IleasnGly 439
QY      878  TAAATCTTTTATATATATATATCTATGAGATGTCAGACAAATATAC---ACTCTGGAA 934
Db      440  TyrasnPhePheSerThrPheProIleGlyCys--IleSerPheasnLeuThrIleThrGlyA 459
QY      935  GAATGGACGTTTGTGGTTATTAGGTAAATTTCTAGTAAACACAGCTGCTGTTTCAG 994
Db      459  laserGlyAlaPhePheTrpThrIleAla---TyrThrSerTyrThrGuaIalaLeuValGlnV 478
QY      995  TT---AACACT 1002
Db      478  aIGluasnThr 481

RESULT 5
US-08-591-685-9
Sequence 9, Application US/08591685
Patent No. 6083733
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Thermostable xylanases
NUMBER OF SEQUENCES: 13
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/591,685
FILING DATE:
CLASSIFICATION: 435
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 544 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-591-685-9

Alignment Scores:
Pred. No.: 0.0472 length: 544
Score: 86.00 Matches: 32
Percent Similarity: 38.28 Conservative: 17
Best Local Similarity: 25.00% Mismatches: 50
Query Match: 4.27% Indels: 29
DB: Gaps: 4

US-09-763-902B-15 (1-1146) x US-08-591-685-9 (1-544)
QY      256  GGCACCTAACACAGGCGTATCTGGACAGGCCACCTTAATATAGGAATGCAATGCTGTTTC 315
Db      281  GyltLeuIleasPglYValGlyLeuGlnGlyHisIleasnValasp-SerProAlaVally 3000
QY      316  AGGCGTGGATGCCCTAAATATATTTTTCACATGATGCCCTCTAGAGAGTTCGAGTCACA 3745
Db      300  sgluIleGluasprThrIleasnLeuPheSerThrIleProGlyLeuGlnIleasnIleTh 3200
QY      376  TATGTATGATTCGCCAACACCAAGTTCATGAGCAGCAACACTCAGTCCCAACTACAGTTTGT 435
Db      320  rgluLeuasp-----IleSerVa 326
QY      436  TGTGACCCAGTGAACCTGTGAAATTTGATGGAAACAACAACACTTC-----481
Db      326  lTyrThrSerSerThrGlnGlnTyrAspThrLeuProGlnaspIleMetIleLysGlnal 346
QY      482  -----TTCAACACGAACCTTCCTGCTGACCTGCTCAGTCCACT-----CCCAACAA 525
Db      346  aleuLysPheLysgluLeuPheGluMetLeuLysAsnGHisSerAspArgIleThrAsnVa 366
QY      526  TACTGTGTGGAGAAATTCGACAGTGAATGCTCCGTTTTCAGAGATTGGTCTAGTAGTTAAAG 585

```

```

Db      366  1ThnleutrlpgleuLysAspAsp-----TyrProTrlPLeuSerLysAspAr 382
Oy      586  GGGCAAAAGTCATTCATTTCATT 607
          1  111
          1  111
Db      382  gSerAsnTrpLeuLeuPhe 389

RESULT 6
PCT-US93-04384-44
; Sequence 44, Application PC/RTUS9304384
; GENERAL INFORMATION:
; APPLICANT: Miller, Timothy J.
; APPLICANT: Kieffer, Sharon
; APPLICANT: Reed, Albert Paul
; APPLICANT: Jones, Elaine V.
; TITLE OF INVENTION: Compositions and Methods for Vaccination
; TITLE OF INVENTION: Against Coronaviruses
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSER: SmithKline Beecham Corporation - Corporate
; ADDRESSEE: Patents
; STREET: 709 Swedeland Road
; CITY: King of Prussia
; STATE: PA
; COUNTRY: USA
; ZIP: 19406-2799
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/04384
; FILING DATE: 19930507
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/882,171
; FILING DATE: 08-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/698,927
; FILING DATE: 13-MAY-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/613,066
; FILING DATE: 14-NOV-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Schreck, Patricia A.
; REGISTRATION NUMBER: 33,777
; REFERENCE/DOCKET NUMBER: SBC H8509-1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 270-5015
; TELEFAX: (215) 270-5090
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1454 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: unknown
; MOLECULE TYPE: protein
; PCT-US93-04384-44

Alignment Scores:
Pred. No.: 0.0784 Length: 1454
Score: 86.00 Matches: 72
Percent Similarity: 35.20% Conservative: 35
Best Local Similarity: 23.68% Mismatches: 109
Query Match: 4.27% Indels: 88
DB: 5. Gaps: 19

US-09-763-902B-15 (1-1146) x PCT-US93-04384-44 (1-1454)
Oy      260  CTAACGACGGCTATCTGAGACAGCGACCTTAATATGGAATGAATGCTGTTTCAGGG 319
          1  111111 111111 111 111111
Db      211  ValThrLeuLeuTyrSerArgSerSerThrAlaThrTrpCyluTyrSerAlaAlaTyrAla 230

```

```

OY 320 CTGGATGCCCTAAATAATTTTGTGACATTCGCTTGTAGTCAAGTCCAGGCAATATG 379
Db 231 TylGInglyValSerAsnPh-----ThrTylTylLysLeuAsn 244
OY 380 TTAGAT-----TGCACACCAAGTTCATGAGCAAGCAACTCACTCC 418
Db 245 ThrAsnGlyLeuLysThrTyrGlnLeuGlyGlnAsp---TyrGlnHisCysThrGlyTyr 263
OY 419 CAATACACAGTCTCTGTGTGACACAGTGCAGTGCATGGAAGTTGATGAAACAAACAACAT 478
Db 264 AlaThrAsnValAlaPheAlaProThrSerGlyGlyTyrIleProAspGly-----Phe 280
OY 479 TTCTTACACCAAGCACTCTCTGTGATCTGCTCACTGCTCACT 517
Db 281 SerPheAsnAsnThrPheLeuLeuThrAsnSerSerThrPheValSerGlyAlaGlyPheVal 300
OY 518 CCCAAC-----AATCTGATGTGGAAGATGCAAGT 547
Db 301 ThrAsnGlnProLeuLeuLeuAsnGlyLeuThrProValProSerPheGlyValAlaAla 320
OY 548 -----GATGCTTCGCGTTTCAGATTTGCTCTACTACTGTAAGGGCAAGTCCATTC 601
Db 321 GlnGlnPheCysPheGlnGlyAlaGlnPheSerGlnCys-----SerGlyVal 336
OY 602 TCATTTGGTCGTCATTACTTCACACATTCGAAATTTATGATATAT 646
Db 337 SerLeuAsnAsnThrValAlaValIleAlaGlyPheAsnLeuAsnPheThrAlaAspValGln 356
OY 647 -----TTGATGTTAGACAGCT 667
Db 357 SerGlyMetCylValAlaThrValPheSerLeuAsnThrThrGlyGlyValIleLeuGlnVal 376
OY 668 ATATGT-----GCTGAACATAAATTTCTTAATATTTTCTATTCTGTCA 712
Db 377 SerCysTyrAsnAspThrValSerGlnSerSerPhe-----TyrSerTyrGlyGln 393
OY 713 GCACCTTTTCTAGCAGCTGCCAGTTTGAGACATTCCTCTTA-----AGAGCT 766
Db 394 IleProPheGlyIleThrAspGlyProAlaGlyTyrCysTyrValIleuTyrAsnGlyThrAla 413
OY 761 TTAATACTATTTTATACATGCCTATATATACATTCACATAAGACATTCATATATATA 820
Db 414 LeuLysTyrLeuGlyThrLeuPro-----ProSerValLysGlnIleAlaIle----- 428
OY 821 TTAACACATGATCTTGCTACTACATACTCACTGACACCAAGCAGCTATTCGAAATAA 880
Db 430 -----SerLysThrGlyHisPheTyr-----IleAsnGlyTyr 440
OY 881 AATCTTTTATATATATATATATATATGATGATGTCAGCAATATATAC---ACTGTGGAGAA 937
Db 441 AsnPheSerThrPheProIleAspCys---IleSerPheAsnLeuThrThrLysAsp 460
OY 938 GTGAGATTTTGTGATTATAGTGAATTTTCTAGTAAACACAGTGCCTGTTTCAGTT- 996
Db 460 ergLysAlaMetThrPheIleAla---TyrThrSerTyrThrGlnAlaLeuValGlnValG 479
OY 997 --AACT 1002
Db 479 LysAsnThr 481

RESULT 7
PCT-US93-04384-45
; Sequence 45, Application: PC/TUS9304384
; GENERAL INFORMATION:
; APPLICANT: Miller, Timothy J.
; APPLICANT: Klepfer, Sharon
; APPLICANT: Reed, Albert P.
; APPLICANT: Jones, Elaine V.
; TITLE OF INVENTION: Compositions and Methods for Vaccination
; TITLE OF INVENTION: Against CoronaviLuses
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SmithKline Beecham Corporation - Corporate

```

ADDRESS: Patents  
STREET: 709 Swedeland Road  
CITY: King of Prussia  
STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/04384  
FILING DATE: 19930507  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/882,171  
FILING DATE: 08-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/698,927  
FILING DATE: 13-MAY-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/613,066  
FILING DATE: 14-NOV-1990  
ATTORNEY/AGENT INFORMATION:  
NAME: Schreck, Patricia A.  
REGISTRATION NUMBER: 33,777  
REFERENCE/DOCKET NUMBER: SGC H85009-1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (215) 270-5015  
TELEFAX: (215) 270-5090  
INFORMATION FOR SEQ ID NO: 45:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1454 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: unknown  
MOLECULE TYPE: protein  
PCT-US93-04384-45

Alignment Scores:			
Pred. No.:	0.0784	Length:	1454
Score:	86.00	Matches:	72
Percent Similarity:	35.208	Conservative:	35
Best Local Similarity:	23.688	Mismatches:	109
Query Match:	4.278	Indels:	88
DB:	5	Gaps:	19

US-09-763-902B-15 (1-1146) x PCT-US93-04384-45 (1-1454)

QY	260	CTAACGAGCGTGTATCTGGACAGGCCACCTTAATAATGAAATGAATGCTGTTTCAGGG	319
Db	211	ValThrLeuLeuLeuTyrSerArgSerSerThrAlaThrTrpGluTyrSerAlaAlaTyrAla	230
QY	320	CTGATGCGCCCTAAATATTTTTCGACACATTGCCCTTAGTCAGTCAGTCCAGGTCATATG	379
Db	231	TyrGlnGlyValSerAspPhe-----ThrTyrTyrIysLeuAsnAsn	244
QY	380	TTAGAT-----TSCAACACAGTTCATGAGCAGACAGCACTCAGTCC	418
Db	245	ThraSngIyLeuLlyThrTyrGlulEuCySgluAsp---TyrGlulnIscyStrHrgIyTyr	263
QY	419	CAAACTACAGTTCCTGTGTGACACAGTGGAACTGTGAAGTTGATGGAACAACAACAT	478
Db	264	AlaThrAsnValAlaPheAlaProThrSerGlyTyrIleProAspIy-----Phe	280
QY	479	TTCTTCACACGAACTCTCTGCTGACTGCTCAGTCACT-----	517
Db	281	SerPheAsnAsnTrpPheLeuLeuThrAsnSerSerThrPheValSerGlyArgPheVal	300
QY	518	CCCAAC-----AATACTGTCGAGATTCGACAGT-----	547
Db	301	ThrAsnGlnProLeuLeuIleAsnIcySngIyProValProSerPheGlyValAlaAla	320



Db 430 -----SerLysTrpGlyHisPheTyr-----IleAsnGly 439  
 QY 878 TAAATCTTTTATATATATATCTATGGATGACGACATATATAC---ACTCTGGGAA 934  
 Db 440 TyrAsnPheserThrPheProIleGlyCys--IleSerPheAsnLeuThrGly 459  
 QY 935 GAATGAGATTGTTGGTTATTAGTTATTTCTAGTAAACAGCTGCTGTTTCAG 994  
 Db 459 alSerGlyAlaPheTrpThrIleAla--TyrThrSerTyrThrGluAlaLeuValGlnV 478  
 QY 995 TT---AACACT 1002  
 Db 478 alGluAsnThr 481  
 RESULT 9  
 PCT-US91-08525-24  
 Sequence 24, Application PC/TUS9108525  
 GENERAL INFORMATION:  
 APPLICANT: SmithKline Beecham, Corporation  
 TITLE OF INVENTION: Recombinant Feline Coronavirus S  
 NUMBER OF SEQUENCES: 54  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SmithKline Beecham Corporation  
 STREET: 709 Swedeland Road  
 CITY: King of Prussia  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19406-2799  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: PCT/US91/08525  
 FILING DATE: 19911114  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/698,927  
 FILING DATE: 13-MAY-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/613,066  
 FILING DATE: 14-NOV-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: King, William T.  
 REGISTRATION NUMBER: 30,954  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 270-5015  
 TELEFAX: (215) 270-5090  
 INFORMATION FOR SEQ ID NO: 24:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 748 amino acids  
 TYPE: AMINO ACID  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 PCT-US91-08525-24  
 Alignment Scores:  
 Pred. No.: 0.0749 Length: 748  
 Score: 85.00 Matches: 69  
 Percent Similarity: 33.77% Conservative: 34  
 Best Local Similarity: 22.62% Mismatches: 112  
 Query Match: 4.22% Indels: 90  
 Gaps: 17  
 US-09-763-902b-15 (1-1146) x PCT-US91-08525-24 (1-748)  
 QY 260 CTACACGAGCTGATCTGACACAGCCACTTATATGAAATGGAATGCTGTTTCAGG 319  
 Db 211 ValThrLeuLeuTyrSerArgSerSerThrAlaThrTrpGluTyrSerAlaAlaTyrAla 230

QY 320 CTGATGCCCTTAATATATTTTGGACATTCGCTTCTAGTGTCCAGCTCAATATG 379  
 Db 231 TyrGlnGlyValSerAsnPhe-----ThrTyrTyrLeuAsnAsn 244  
 QY 380 TTACAT-----TGCCACACAGTTCATGTAGCAACACACAGTC 418  
 Db 245 ThrAsnGlyLeuLysThrTyrGluLeuGlyGlnPsp---TyrGluHisCysTyrGlyTyr 263  
 QY 419 CAACACAGCTTCTGTTGTTGACAGCTGAGACTGGAAGTTTGATGGAAACAAACACAT 478  
 Db 264 AlaThrAsnValPheAlaProThrSerGlyLysTyrIleProAspGly-----Phe 280  
 QY 479 TTCTTCACACAGAACTTCCTGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 517  
 Db 281 SerPheAsnAsnTrpPheLeuLeuThrAsnSerSerThrPheValSerGlyArgPheVal 300  
 QY 518 CCCAAC-----AATACGTGTGGAAGATTGCAACT----- 547  
 Db 301 ThrAsnGlnProLeuLeuIleAsnCysLeuTrpProValProSerPheGlyValAlaAla 320  
 QY 548 -----GATTCCTTCCTGTTTCAAGATTGCTCTAGTACTTAAGGGGCAAAAGTCATTC 601  
 Db 321 GlnGluPheCysPheGlnGlyAlaGlnPheSerGlnCys-----AsnGlyVal 336  
 QY 602 TCATTTGGTCCATTAGTTCAGCAATTCGAATTTATGGAATATAT----- 646  
 Db 337 SerLeuAsnAsnThrValAlaPspValIleArgPheAsnLeuAsnPheThrAlaPspValGln 356  
 QY 647 -----TTT 649  
 Db 357 SerGlyMetGlyAlaThrValPheSerLeuAsnThrThrGlyGlyValIleLeuGluIle 376  
 QY 650 GATGTGAGACCATATATATGTCGCAACTAATTTCTTATATTTCTATCTGCTG 709  
 Db 377 SerCysTyrSerAspThrValSerGlnSerSerSer-----TyrSerTyrGly 392  
 QY 710 TCACACACCTTTCTAGACAGTCGCCAGTTTGAGCATTCGCCCTCA-----AGA 757  
 Db 393 GluIleProPheGlyIleThrAspGlyProArgTyrCysTyrValLeuTyrAsnGlyThr 412  
 QY 758 GCTTAAACATATTTTTCATACGCTTATATACATTCCTAATGACATTCCTTATATA 817  
 Db 413 AlaLeuLysTyrLeuGlyThrLeuPro-----ProSerValLysGluIleAlaIle 429  
 QY 818 ATATTAAACATATATCTGTGCTACTACATCTACCTGACACCCAGCTTTGCCAANA 877  
 Db 430 -----SerLysTrpGlyHisPheTyr-----IleAsnGly 439  
 QY 878 TAAATCTTTTATATATATATCTATGGATGACGACATATATAC---ACTCTGGGAA 934  
 Db 440 TyrAsnPheserThrPheProIleGlyCys--IleSerPheAsnLeuThrGly 459  
 QY 935 GAATGAGATTGTTGGTTATTAGTTATTTCTAGTAAACAGCTGCTGTTTCAG 994  
 Db 459 alSerGlyAlaPheTrpThrIleAla--TyrThrSerTyrThrGluAlaLeuValGlnV 478  
 QY 995 TT---AACACT 1002  
 Db 478 alGluAsnThr 481  
 RESULT 10  
 PCT-US93-04384-4  
 Sequence 4, Application PC/TUS9304384  
 GENERAL INFORMATION:  
 APPLICANT: Miller, Timothy J.  
 APPLICANT: Klepfer, Sharon  
 APPLICANT: Reed, Albert Paul  
 APPLICANT: Jones, Elaine V.  
 TITLE OF INVENTION: Compositions and Methods for Vaccination  
 NUMBER OF SEQUENCES: 48  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: SmithKline Beecham Corporation - Corporate



```

ADDRESS: Patents
ADDRESS: 709 Swedeland Road
CITY: King of Prussia
STATE: PA
COUNTRY: USA
ZIP: 19406-2799
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/04384
FILING DATE: 19930507
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/882,171
FILING DATE: 08-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/698,927
FILING DATE: 13-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/613,066
FILING DATE: 14-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Schreck, Patricia A.
REGISTRATION NUMBER: 33,777
REFERENCE/DOCKET NUMBER: SPC H85009-1
TELECOMMUNICATIONS INFORMATION:
TELEPHONE: (215) 270-5015
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 748 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-04384-4.

Alignment Scores:
Pred. NO: 0.0749 Length: 748
Score: 85.00 Matches: 69
Percent Similarity: 33.77% Conserved: 34
Best local Similarity: 22.62% Mismatches: 112
Query Match: 4.22% Indels: 90
DB: 5 Gaps: 17

US-09-763-902B-15 (1-1146) x PCT-US93-04384-4 (1-748)
QY 260 CTAACGAGCGTGTATCTGACAAAGCCACCTTAATGATGAAATGCTGTTGAGCG 319
DB ::::||||| ::::||||| ::::|||||
QY 211 VAIHTIRLeuTyrSerIarGserSerThrAlaThrIpsLuTyrSerAlaIaTyrAla 230
DB :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 320 CTGATGCCCTTAATTAATTTTTTTTGACACATTGCCCTTAGTGAGTCCAGGTCAATATG 379
DB ::::||||| ::::||||| ::::||||| ::::||||| ::::|||||
QY 221 TyrGInGlyValSerAsnPh-----ThyTyrTyrIysLeuAsnAsn 244
DB ::::||||| ::::||||| ::::||||| ::::||||| ::::|||||
QY 380 TTAGAT-----TGCCAAACGATTGATGAGCAGCAGCACTGAC 418
DB ::::||||| ::::||||| ::::||||| ::::||||| ::::|||||
QY 245 ThrAsnGlyLeuLysThrTyrIguLeuGlyAsp--TyrIguIscYsthrGlyTyr 263
DB :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 419 CAAATGACGTTCTGTTGTGACACGAGTGAAGCTGTAAGTTGATGGAACAACAACAT 478
DB :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 264 AlaThrAsnValAlaPheAlaProThrSerGlyTyrIleProAspGly-----Phe 280
DB :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 479 TTCCTTCAACGAACTCTGCTGACCTGCTCAGTCCACT----- 517
DB :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 281 SerPheAsnAsnTyrPheLeuLeuThrAsnSerSerThrPheValSerGlyArgPheVal 300
DB :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 518 CCCAAC-----AAATGCTGTGGAAGATGCAAGT----- 547
DB :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
QY 301 ThrAsnGlnProLeuLeuIleAsnGlySerLeuTyrProValProSerPheGlyValAlaIa 320
DB :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||

```

```

0Y 548 -----GATTGCTTCGTTTCAAGATGGCTGTAGTAGTTAAAGGGCAAAAGTCATTC 601
0Y 321 GInglupheCsPhnglUGlYlaGlnPheSerGInCs-----AsnGlyVal 336
0Y 602 TCATTGGTCCATTAGTTCACAGCAATGGAATTTATGTAATAT- 646
0Y 337 SerLeuAsnAsnThrValAspValIleArgPheAsnLeuAsnPheThrAlaAspValGln 356
0Y 647 -----TTT 649
Db 357 SerGlyMetGlyAlaThrValPheSerLeuAsnThrThrglyGlyValIleLeuGluIle 376
0Y 650 GATTGTAGAGACGACATATATATGCTGCTGAACATAATTCCTTTAATTTTCATTCCTG 709
Db 377 SerCysTyrSerAspThrValSerGlnUserSer-----TyrSerTyrGly 392
0Y 710 TCAGACACCTTTTCTAGACAGCTCCAGTTTGGAGACATGGCCCTCA-----AGA 757
Db 393 GluIleProPheGlyIleThrAspGlyProArgTyrCysTyrValLeuTyrAsnGlyThr 412
0Y 758 GCTTTAAACATATTTTTCATACGCTTATATACATTCACATGATGACATTCCTTATATA 817
Db 413 AlaLeuTyrLeuGlyThrLeuPro-----ProSerValIleGluIleAlaIle--- 429
0Y 818 AATTTAAACATGATCTGTGTACTACATPACTGACGTGAACCCAGCCTATGCAAAA 877
Db 430 -----SerIstyrpGlyHisPheTyr-----IleAsnGly 439
0Y 878 TAAATCTTTTATATATATATCTATGATGGATGCACACAAATATAC---ACTCTGGGA 934
Db 440 TyrAsnPhnPheSerThrPheProIleGlyCys--IleSerPheAsnLeuThrThrgly 459
0Y 935 GAAGTGGACCTTTTGGTTATTAGTTAAATTTTCTAGTAACAAACAGCTGGCTTTTCAG 994
Db 459 alSerGlyAlaPheThrPheThrIleAla---TyrThrSerTyrThnGluAlaLeuValGln 478
0Y 995 TT---AACACT 1002
Db 478 alGluAspThr 481

RESULT 11
US-08-915-795-8
Sequence 8, Application US/08915795
Patent No. 6235713
GENERAL INFORMATION:
APPLICANT: Marc G. ACHEN
APPLICANT: Andrew F. MILKS
APPLICANT: Steven A. STACKER
APPLICANT: Kari ALITALO
TITLE OF INVENTION: GROWTH FACTOR
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: Evenson, McKeown, Edwards & Lenahan P.L.L.C.
STREET: 1200 G Street, NW, Suite 700
CITY: Washington
STATE: DC
COUNTRY: United States of America
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/915,795
FILING DATE:
CLASSIFICATION: 536
ATTORNEY/AGENT INFORMATION:
NAME: EVANS, Joseph D.
REGISTRATION NUMBER: 26,269
REFERENCE/DOCKET NUMBER: 1064/42983
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 628-8800

```



```

Db 297 GlyArgPheValThrAsnGlnProLeuValAsnCysLeuThrProValProSerPhe 316
QY 548 -----GATTCCTCCCTTTTCAAGATGGCTCTGACTTAAAGGGC 589
Db 317 GlyValAlaIaGlnGluPheCysPheGluGlyAlaGlnPheSerIncys 333
QY 590 AAAAGTCATTCCTCATTTGGTCATTAAGTCCAGCAATGAATTAATGATTAAT 646
Db 334 ---AsnGlyValSerLeuAsnAsnThrValAspValIleArgPheAsnLeuAsnPheThr 352
QY 646 ----- 646
Db 353 AlaAspValGlnSerGlyMetGlyAlaThrValPheSerLeuAsnThrThrGlyVal 372
QY 647 -----TTTGATGTAGAGAGCACTAATATATGCTGAAACTAAATTTCTTAATAT 697
Db 373 IleLeuGluIleSerCysTyrSerAspThrValSerGlyUserSerSer 388
QY 698 TTTCTATTCTGTCAGACACCTTTCTAGACAGCTCCAGTTGAGCATTCCTCTA 754
Db 389 TyrSerTyrGlyGluIleProPheGlyIleThrAspGlyProArgTyrCysTyrValLeu 408
QY 755 -----AGACCTTAAACATATTTTATACATGCTATATACCTTCCACTAATGAC 805
Db 409 TyrAsnGlyThrAlaLeuIleGlyThrLeuPro-----ProSerValGlyGlu 426
QY 806 ATCTCTATATATATATATTAACACATGATCTGTGACTACATCACTCACTGAAACCCAGC 865
Db 427 IleAlaIle-----SerTyrTrpGlyHisPheTyr----- 436
QY 866 CTATTCGCAAAAAATCTTTTATATATATATCTGTGAGCTGACCAATATATAC 924
Db 437 ---IleAsnGlyTyrAsnPhePheSerThrPheProIleGlyCys--IleSerPheAsn 455
QY 925 --ACTCGGAGAGAGAGAGTGTGTTAGTAAATTTCTAGTAAACACAGCT 982
Db 455 eutThrGlyAlaSerGlyAlaPheThrPheIleAla---TyrThrSerTyrThrGlu 474
QY 983 GCCTGTTTCAGTT--ACACT 1002
Db 474 IalLeuValGlnValGluAsnThr 481

```

RESULT 13  
PCT-US93-04384-46  
Sequence 46, Application PC/TUS9304384  
GENERAL INFORMATION:

APPLICANT: Miller, Timothy J.  
APPLICANT: Klepfer, Sharon  
APPLICANT: Reed, Albert Paul  
APPLICANT: Jones, Elaine V.  
TITLE OF INVENTION: Compositions and Methods for Vaccination  
TITLE OF INVENTION: Against Coronaviruses  
NUMBER OF SEQUENCES: 48  
CORRESPONDENCE ADDRESS:

ADDRESSEE: SmithKline Beecham Corporation - Corporate  
ADDRESSEE: Patents  
STREET: 709 Swedeland Road  
CITY: King of Prussia

STATE: PA  
COUNTRY: USA  
ZIP: 19406-2799

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US93/04384  
FILING DATE: 19930507

CLASSIFICATION:

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/882,171

```

FILING DATE: 08-MAY-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/698,927
FILING DATE: 13-MAY-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/613,066
FILING DATE: 14-NOV-1990
ATTORNEY/AGENT INFORMATION:
NAME: Schreck, Patricia A.
REGISTRATION NUMBER: 33,777
REFERENCE/DOCKET NUMBER: SBC H85009-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 270-5015
TELEFAX: (215) 270-5090
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 1454 amino acids
TYPE: AMINO ACID
TOPOLOGY: unknown
MOLECULE TYPE: protein
PCT-US93-04384-46

Alignment Scores:
Pred. No.: 0.403 Length: 1454
Score: 80.50 Matches: 74
Percent Similarity: 32.83% Conservative: 34
Best Local Similarity: 22.49% Mismatches: 124
Query Match: 4.00% Indels: 97
DB: 5 Gaps: 18

US-09-763-902b-15 (1-1146) x PCT-US93-04384-46 (1-1454)
QY 209 GAGAGGTTGTCATATTTAC-----TATGAGACAAATGATATA 247
Db 187 AspIlePheValThrAlaTyrIleSerGlyArgSerTyrAsnThrPheAsnAsn 206
QY 248 AGAAGAGGGGAGCTACACAGGCTGTATCTGACACAGCCACTTAATATGAAATGAAAT 307
Db 207 TrpPheAsnAsnValThrLeuLeuTyrSerArgSerThrAlaThrTrpGluTyrSer 226
QY 308 GCTGTTTCAGGGCTGATGCTTAAATATTT-----TTTGCAC 346
Db 227 AlaAlaTyrValTyrGlnGlyValSerAsnPheThrTyrTyrTyrLeuAsnAsnThrAsn 246
QY 347 ACATTCGCTTCTAGTGAATTCAGTCAATATGTTAGTTGCCAACCAGTTGATGACCA 406
Db 247 GlyLeuTyrThrTyrGluPhe-----CysGluAsp--TyrGluTyr 259
QY 407 GCAACTGAGTCCCAACATACAGTCTGTGTGACACAGTGAAGTGAAGTTGATGCA 466
Db 260 CysThrGlyTyrAlaThrAsnValPheAlaProThrValGlyTyrIleProAspGly 279
QY 467 AACAAACAACATTTCTTACACAGACTTCCTGCTGACTGCTCACTCAGT 517
Db 280 -----PheSerPheAsnAsnThrPheLeuLeuThrAsnSerSerThrPheValSer 296
QY 518 -----CCCAAC-----AATATCTGTGTGAAGATTCAGTCACT 547
Db 297 GlyArgPheValThrAsnGlnProLeuValAlaCysLeuThrProValProSerPhe 316
QY 548 -----GATTCCTCCGTTTTCAGAGATGCTGCTAGTAAAGGGGC 589
Db 317 GlyValAlaIaGlnGluPheCysPheGluGlyAlaGlnPheSerIncys 333
QY 590 AAAAGTCATTCCTCATTTGGTCATTAAGTCCAGCAATGAATTAATGATTAAT 646
Db 334 ---AsnGlyValSerLeuAsnAsnThrValAspValIleArgPheAsnLeuAsnPheThr 352
QY 646 ----- 646
Db 353 AlaAspValGlnSerGlyMetGlyAlaThrValPheSerLeuAsnThrThrGlyVal 372
QY 647 -----TTTGATGTAGAGAGCACTAATATATGCTGAAACTAAATTTCTTAATAT 697

```

APPLICATION NUMBER: US 07/698,927

RESULT 15  
 US-08-392-459-32  
 Sequence 32, Application US/08392459  
 Patent No. 6280974  
 GENERAL INFORMATION:  
 APPLICANT: SmithKline Beecham, Corporation  
 TITLE OF INVENTION: Recombinant Feline Coronavirus S  
 TITLE OF INVENTION: Proteins  
 NUMBER OF SEQUENCES: 54  
 CORRESPONDENCE ADDRESSES:  
 ADDRESSEE: SmithKline Beecham Corporation  
 STREET: 709 Swedeland Road  
 CITY: King of Prussia  
 STATE: PA  
 COUNTRY: USA  
 ZIP: 19406-2799  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/392,459  
 FILING DATE:  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/698,927

FILING DATE: 13-MAY-1991  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 07/613,066  
 FILING DATE: 14-NOV-1990  
 ATTORNEY/AGENT INFORMATION:  
 NAME: King, William T.  
 REGISTRATION NUMBER: 30,954  
 REFERENCE/DOCKET NUMBER: SBC 14532B  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (215) 270-5015  
 TELEFAX: (215) 270-5090  
 INFORMATION FOR SEQ ID NO: 32:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1454 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-392-459-32

Alignment Scores:  
 Pred. No.: 0.543 Length: 1454  
 Score: 79.50 Matches: 77  
 Percent Similarity: 34.15% Conservative: 35  
 Best Local Similarity: 23.48% Mismatches: 121  
 Query Match: 3.95% Indels: 95  
 DB: 4 Gaps: 20

US-09-763-902b-15 (1-1146) x US-08-392-459-32 (1-1454)

```

OY 209 GAGGAGTTTGTCATATTTC-----TATGAGACAATGATATA 247
Db 187 AspIuPhValIthAlaIyrlIleSerGlyArgSerTyraSnItrPasnIleAsnAsn 206
OY 248 AGAAGAGGCGCATTACAGCCTGATCTGACAGGCGCATTAAATGGAATGAAAT 307
Db 207 TrpPheAsnAsnValIthLeuLeuTySerArgSerSerIthAlaIthIrpGluTySer 226
OY 308 GCTGCTTCAGGCGCGATGCCCTAAATATTT-----TTTGAC 346
Db 227 AlaIaIyryValIyrgInglyValSerAsnPhethIryTyrlYsLeuAsnAsnIthrasn 246
OY 347 ACATTGCTCTAGTTCAGTTCAGGTCGAATATGTTAGATTGCCAACCAGTTCATGAGCAA 406
Db 247 GlyLeuIystrIyrgIuPhe-----CysGluAsp---TyrgIuTy 259
OY 407 GCAACTGAGTCCCAACTACAGTTCTGTGTGACCACTGGAAGTGAAGTTGATGA 466
Db 260 CysThrGlyTyraIethIthAsnValPheAlaIroThryValGlyIyrlIleProAspIy 279
OY 467 AACAAACACATTTCTTCACAGCACTCTGCTGACTGCTCAGTCCACT----- 517
Db 280 -----PheSerPheAsnAsnItrPheLeuLeuIthrasnSerSerIthPheValSer 296
OY 518 -----CCCAC-----AATACGTGTGGAAGATTGCACT--- 547
Db 297 GlyArgPheValIthrasnGlnProLeuLeuValAsnCysLeuItrProValProSerPhe 316
OY 548 -----GATTGCTCCGTTTCAAGATTGCTAGTAGTAAAGGCGC 589
Db 317 GlyValAlaIaIaGlnGluPheCysPheGluGlyAlaGlnPheSerGlnCys----- 333
OY 590 AAAAGTCCATTCTGATTGCTCATTTAGTTCAGCAATGGAATTTATGTAATAT 646
Db 334 ---SerGlyValSerLeuAsnAsnIthValAspValIleArgPheAsnLeuAsnPhethr 352
OY 647 -----TTGATGTAGAGCACT----- 664
Db 353 AlaAspValGlnSerGlyMetGlyAlaIthValPheSerLeuAsnIthIthGlyVal 372
OY 665 -----ATAATATGT-----GCTGAACATAATTTCTTAATATTT 700
Db 373 IleLeuGluValSerCysTyraAsnAspIthValSerGluSerSerPhe-----Tyr 389
  
```

```

OY 701 CTATTCCTGTCAGCACCTTTCTAGACAGCTGCCAGTTTGAGACATGGCCCTCTA----- 754
Db 390 SerIyrgIyGluIleProPheGlyIleIthrasnIyProArgTyrcysTyryValLeuTy 409
OY 755 -----AGAGCTTAAACTATTTTTCATGCTTATATATATATATATATATATATAT 808
Db 410 AsnGlyIthAlaLeuIystrIyLeuGlyIthLeuPro-----ProSerValIyGluIle 427
OY 809 CTTATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 868
Db 428 AlaIle-----SerIystrIyGlyIthIlePheTy----- 436
OY 869 TTGCAAAAATAAATCTTTTATATATATATATATATATATATATATATATATATATATAT 925
Db 437 IleAsnGlyTyraSnPhePheSerIthPheProIleAspCys---IleSerPheAsnLeu 456
OY 926 CTCGTGGAGAAGTGAAGTATTTTGTATTATGTTATATATATATATATATATATATATAT 985
Db 456 hrIthGlyAspSerGlyAlaPheItrPthIleAla---TyrlThSerTyrlthGluAla 475
OY 986 TGTTTTCAGTT---AACACT 1002
Db 475 euValGlnValGluAsnIth 481
  
```

Search completed: June 24, 2003, 19:40:32  
 Job time : 32.789 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Comphen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: June 24, 2003, 19:20:48 ; Search time 54,8062 Seconds

(without alignments)  
5572.555 Million cell updates/sec

Title: US-09-763-902b-15

Perfect score: 2013

Sequence: 1 taccctcgsgagagaatg99.....tttaagratattttgtttt 1146

Scoring table:

BLOSUM62	
Xgapop 10.0, Xgapext 0.5	
Ygapop 10.0, Ygapext 0.5	
Fgapop 6.0, Fgapext 7.0	
Delop 6.0, Delext 7.0	

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame-n2p.model -DEV=xlp  
-Q/cg2.1/USPTO-SPOOL/US09763902/runat\_24062003.130340.12405/app\_query.fasta.1.2254  
-DB-A.Geneseq.101002 -OPMT=faatan -SUFFIX=02p.rag -MINMATCH=0.1 -LOOPTOL=0  
-LOOPTXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=bloms62 -TRANS=humand0.cdl  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTPM=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09763902 -ECGN\_L1=1.162 -ecgnat\_24062003.130340.12405 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=30 -WAIT -DSPBLOCK=100 -IONLOCK  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A.Geneseq.101002.\*  
1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.\*  
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.\*

Prod. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	874	43.4	197	22	AB95379
2	749	37.2	142	21	AA982323
3	749	37.2	142	23	AB804459
4	727	36.1	162	22	AA016227
5	584.5	29.0	164	22	AA016228
6	584	29.0	140	21	AA982318
7	310	15.4	133	22	AB865017
8	114	5.7	688	22	AB867404
9	114	5.7	680	22	AB866052
10	113.5	5.6	125	21	AA012073
11	113.5	5.6	153	21	AA005145
12	112.5	5.6	123	21	AA008682
13	112.5	5.6	126	21	AA008681
14	112.5	5.6	131	21	AA005147
15	112.5	5.6	134	21	AA005146
16	105	5.2	122	21	AA029868
17	105	5.2	157	21	AA029867
18	102.5	5.1	130	22	AB871855
19	91.5	4.5	170	22	AA025600
20	90.5	4.5	128	21	AA021068
21	90.5	4.5	456	21	AA021067
22	88.5	4.4	472	21	AA046342
23	88.5	4.4	143	21	AA046342
24	87.5	4.3	460	21	AA046343
25	87.5	4.3	699	14	AA030512
26	87	4.3	1454	13	AA024395
27	87	4.3	1454	14	AA042464
28	87	4.3	1454	14	AA042477
29	86	4.3	471	21	AA020809
30	86	4.3	544	17	AA087014
31	86	4.3	703	22	AB013036
32	86	4.3	1017	22	AB008678
33	86	4.3	1454	14	AA042473
34	85	4.2	1594	9	AA081183
35	84.5	4.2	459	21	AB020810
36	84	4.2	938	22	AB010442
37	84	4.1	1127	22	AA033877
38	83.5	4.1	1158	22	AA036929
39	83.5	4.1	284	20	AA035356
40	83	4.1	746	23	AA074631
41	82	4.1	752	22	AA070205
42	82	4.1	755	22	AA070206
43	82	4.1	755	22	AB061889
44	82	4.1	755	23	AA051198
45	82	4.1	755	23	AA051198

## ALIGNMENTS

RESULT 1  
ID AAB95379 standard; Protein; 197 AA.  
AC AAB95379;  
XX  
DT 26-JUN-2001 (first entry)  
XX  
DE Human protein sequence SEQ ID NO:17707.  
XX  
KW Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX  
OS Homo sapiens.  
XX  
PN EP1074617-A2.  
XX  
PD 07-FEB-2001.  
XX  
PF 28-JUL-2000; 2000EP-0116126.

XX 29-JUL-1999; 99JP-0248036.  
 PR 27-AUG-1999; 99JP-0300253.  
 PR 11-JAN-2000; 2000JP-0118776.  
 PR 02-MAY-2000; 2000JP-0183767.  
 PR 09-JUN-2000; 2000JP-0241899.  
 XX  
 PA (HELI-) HELIX RES INST.  
 XX  
 PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
 DR WPI: 2001-318749/34..  
 XX  
 PT Primer sets for synthesizing polynucleotides, particularly the 5602  
 PT full-length cDNAs defined in the specification, and for the detection  
 PT and/or diagnosis of the abnormality of the proteins encoded by the  
 PT full-length cDNAs.  
 XX  
 PS Claim 8: SEQ ID 17707; 2537bp + CD ROM; English.  
 XX  
 CC The present invention describes primer sets for synthesizing 5602  
 CC full-length cDNAs defined in the specification. Where a primer set  
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
 CC to the complementary strand of a polynucleotide which comprises one of  
 CC the 5602 nucleotide sequences defined in the specification, where the  
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
 CC of an oligonucleotide comprising a sequence complementary to the  
 CC complementary strand of a polynucleotide which comprises a 5'-end  
 CC sequence and an oligonucleotide comprising a sequence complementary to a  
 CC polynucleotide which comprises a 3'-end sequence, where the  
 CC oligonucleotide comprises at least 15 nucleotides, and the combination of  
 CC the 5'-end sequence/3'-end sequence is selected from those defined in  
 CC the specification. The primer sets can be used in antisense therapy and  
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
 CC particularly full-length cDNAs. The primers are also useful for the  
 CC detection and/or diagnosis of the abnormality of the proteins encoded by  
 CC the full-length cDNAs. The primers allow obtaining of the full-length  
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and  
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAH92446 to  
 CC AAH5893 represent human amino acid sequences; and AAH13629 to AAH13632  
 CC represent oligonucleotides, all of which are used in the exemplification  
 CC of the present invention.  
 XX  
 SQ Sequence 197 AA:

Alignment Scores:  
 Pred. No.: 4,19e-98 Length: 197  
 Score: 874.00 Matches: 167  
 Percent Similarity: 98.82% Conservative: 0  
 Best Local Similarity: 98.82% Mismatches: 2  
 Query Match: 43.42% Indels: 0  
 DB: 22 Gaps: 0

US-09-763-902B-15 (1-1146) x AAB95379 (1-197)

OY 74 TCATCCCGACCGCCGCGACCGACGAGTAGAGAGTAGAGCCGACACTGCGAGAA 133  
 DB 29 SerHisSerSerProAlaAspArgThrArgGlnGluValThrProThrLeuProGlu 48  
 OY 134 CACACTGCTACAGGTCCAGATGCCACGCTGCTGAGATTAAACTATATGATCAG 193  
 DB 49 HlStrHlathrHrSerGlnMetalarHrSerLeuAspPheLysThrTyrAlaAspGln 68  
 OY 194 GCATGTAGAGCTGAGAGAGTTTGTCAATATTTACTATGACACATGATGAAGAAGA 253  
 DB 69 AlaCysArgAlaAlaGlnGluPheValAsnIleTyrTyrGlnThrMetAspLysArgArg 88  
 OY 254 CGGGCACTACACAGGCTGTATCTGACACAGGCCACCTTAATATGGAAGAAGAACTGCT 313  
 DB 89 ArgAlaLeuThrHrArgLeuTyrLeuAspLysAlaThrIleuIleTyrAsnGlnLysAlaVal 108  
 OY 314 TCAGGCTGATGCCCTAAATATTTTGTGACACATGCTTCTAGAGAGTTCAGAGTC 373

DB 109 SerGlyLeuAspAlaLeuAsnAsnPheAspHrLeuProSerSerGlnPheGlnVal 128  
 OY 374 AATATGTAGATTGCGCAACAGCTTCATGACAGACAACTGATCCCAACACTGTTT 433  
 DB 129 AsnMetLeuAspCysGlnProValHisGlnIleAlaThrGlnSerGlnThrValLeu 148  
 OY 434 GTTGTGACCACTGAGCACTGATGATGATGAACAACAACAATTTCTCAACTAGAAC 493  
 DB 149 ValValThrSerGlyThrValLysPheAspLysGlnHisPheAsnGlnAsn 168  
 OY 494 TTCTGTGACTGCTCACTGCTCCACCACTGCTGAGAGATGCAAGTATTC 553  
 DB 169 PheLeuLeuThrAlaGlnSerThrProAsnAsnThrValTyrPlyIleAlaSerAspCys 188  
 OY 554 TTCGTTTCAAGATGTGCTAGAGT 580  
 DB 189 PheArgPheGlnAspTyrPheSerSer 197

# RESULT 2

AAH82323

ID AAH82323 standard; Protein: 142 AA.

XX AAH82323;

AC 19-JUN-2000 (first entry)

DE Human protein transport molecule (PTM) SEQ ID NO:7:

XX Human: protein transport molecule; PTM: diagnosis; cytostatic;  
 KW antiallergic; antidiabetic; immunosuppressant; antiarteriosclerotic;  
 KW antiallergic; antidiabetic; antilipemic; antihypertensive; osteoporotic;  
 KW dermatological; antianemic; antipsoriatic; hepatotropic; antigout;  
 KW antinflammatory; antiHIV; protein transport regulator; cancer;  
 KW immune disorder; cell proliferative disorder; secretory disorder;  
 KW urticaria; allergy; abnormal vesicle trafficking; asthma;  
 KW autoimmune haemolytic anaemia.

OS Homo sapiens.

PN WO200012703-A2.

PD 09-MAR-2000.

XX 26-AUG-1999; 99WO-US19616.

PR 27-AUG-1998; 98US-0098206.

XX (INCY-) INCYTE PHARM INC.

XX Tang YF, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ;

XX Gorgone GA, Baughn MR, Patterson C;

DR WPI: 2000-256642/22.

DR N-PSDB: AAA08041.

PT New human protein transport-associated polypeptide and polynucleotide  
 PT useful for diagnosis, prevention and treatment of cell proliferative  
 PT and secretory disorders such as leukemia, cystic fibrosis

PS Claim 1: Page 68; 75pp; English.

CC AAA08035 to AAA08042 encode the human protein transport-associated  
 CC molecules (PTMs) given in AAH82317 to AAH82324. The PTMs have  
 CC cytotostatic, antiallergic, antianemic, immunosuppressant,  
 CC antiatherosclerotic, antiallergic, antidiabetic, antilipemic,  
 CC antirheumatic, osteoporotic, dermatological, antianemic, antipsoriatic,  
 CC hepatotropic, antigout, antinflammatory and antiHIV activities, and  
 CC regulate protein transport. PTM proteins and antagonists are useful for  
 CC preventing or treating a disorder associated with decreased or increased  
 CC expression or activity of PTM. PTM polynucleotides are useful for  
 CC diagnosing conditions associated with PTM, comprising detecting PTM by  
 CC forming a hybridisation complex, preferably after PCR amplifying the





XX Human novel secreted protein, Seg ID 1180.  
DE  
XX  
KW Human; immunosuppressive; antiarthritic; antirheumatic;  
KW cyostatic; cardiant; vasotropic; cerebroprotective; noctopic;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW vulnerary; secreted protein; rheumatoid arthritis;  
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KW cerebrovascular disorder; cerebral ischaemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KW corneal infection; wound healing; epithelial cell proliferation;  
KW skin ageing; food additive; preservative; antiproliferative.  
XX  
XX Homo sapiens.  
XX  
XX WO20015322-A2.  
XX  
XX  
XX 02-AUG-2001.  
XX  
XX  
XX 17-JAN-2001; 2001WO-US01341.  
XX  
XX 31-JAN-2000; 2000US-0179065.  
XX 04-FEB-2000; 2000US-0180628.  
XX 24-FEB-2000; 2000US-0184664.  
XX 02-MAR-2000; 2000US-0186350.  
XX 16-MAR-2000; 2000US-0189874.  
XX 17-MAR-2000; 2000US-0190076.  
XX 18-APR-2000; 2000US-0198123.  
XX 19-MAY-2000; 2000US-0205515.  
XX 07-JUN-2000; 2000US-0209467.  
XX 28-JUN-2000; 2000US-0214886.  
XX 30-JUN-2000; 2000US-0215135.  
XX 07-JUL-2000; 2000US-0216647.  
XX 07-JUL-2000; 2000US-0216880.  
XX 11-JUL-2000; 2000US-0217487.  
XX 11-JUL-2000; 2000US-0217496.  
XX 14-JUL-2000; 2000US-0218290.  
XX 26-JUL-2000; 2000US-0220963.  
XX 26-JUL-2000; 2000US-0220964.  
XX 14-AUG-2000; 2000US-0224518.  
XX 14-AUG-2000; 2000US-0224519.  
XX 14-AUG-2000; 2000US-0225213.  
XX 14-AUG-2000; 2000US-0225214.  
XX 14-AUG-2000; 2000US-0225265.  
XX 14-AUG-2000; 2000US-0225267.  
XX 14-AUG-2000; 2000US-0225268.  
XX 14-AUG-2000; 2000US-0225270.  
XX 14-AUG-2000; 2000US-0225447.  
XX 14-AUG-2000; 2000US-0225757.  
XX 14-AUG-2000; 2000US-0225758.  
XX 14-AUG-2000; 2000US-0225759.  
XX 18-AUG-2000; 2000US-0226279.  
XX 22-AUG-2000; 2000US-0226681.  
XX 22-AUG-2000; 2000US-0226868.  
XX 22-AUG-2000; 2000US-0227182.  
XX 23-AUG-2000; 2000US-0227009.  
XX 30-AUG-2000; 2000US-0228924.  
XX 01-SEP-2000; 2000US-0229287.  
XX 01-SEP-2000; 2000US-0229343.  
XX 01-SEP-2000; 2000US-0229344.  
XX 01-SEP-2000; 2000US-0229345.  
XX 05-SEP-2000; 2000US-0229509.  
XX 05-SEP-2000; 2000US-0229513.  
XX 06-SEP-2000; 2000US-0230437.  
XX 06-SEP-2000; 2000US-0230438.  
XX 08-SEP-2000; 2000US-0231242.  
XX 08-SEP-2000; 2000US-0231243.  
XX 08-SEP-2000; 2000US-0231244.  
XX 08-SEP-2000; 2000US-0231413.  
XX 08-SEP-2000; 2000US-0231414.  
XX 08-SEP-2000; 2000US-0232080.  
XX 08-SEP-2000; 2000US-0232081.  
XX 12-SEP-2000; 2000US-0231968.

PR 14-SEP-2000; 2000US-0232397.  
PR 14-SEP-2000; 2000US-0232398.  
PR 14-SEP-2000; 2000US-0232399.  
PR 14-SEP-2000; 2000US-0232400.  
PR 14-SEP-2000; 2000US-0232401.  
PR 14-SEP-2000; 2000US-0233063.  
PR 14-SEP-2000; 2000US-0233064.  
PR 21-SEP-2000; 2000US-0233065.  
PR 21-SEP-2000; 2000US-0234223.  
PR 21-SEP-2000; 2000US-0234274.  
PR 25-SEP-2000; 2000US-0234997.  
PR 25-SEP-2000; 2000US-0234998.  
PR 26-SEP-2000; 2000US-0235484.  
PR 27-SEP-2000; 2000US-0235834.  
PR 27-SEP-2000; 2000US-0235836.  
PR 29-SEP-2000; 2000US-0236337.  
PR 29-SEP-2000; 2000US-0236367.  
PR 29-SEP-2000; 2000US-0236368.  
PR 29-SEP-2000; 2000US-0236369.  
PR 29-SEP-2000; 2000US-0236370.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237037.  
PR 02-OCT-2000; 2000US-0237038.  
PR 02-OCT-2000; 2000US-0237039.  
PR 13-OCT-2000; 2000US-0239935.  
PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0244617.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246609.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249264.  
PR 17-NOV-2000; 2000US-0249265.  
PR 17-NOV-2000; 2000US-0249297.  
PR 17-NOV-2000; 2000US-0249299.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.

PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251989.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX Rosen CA, Barash SC, Ruben SM.  
XX WPI; 2001-489783/53.  
DR N-PSDB; AAS26214.  
XX  
PT New nucleic acid molecules encoding 461 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives.  
XX  
PS Claim 11; SEQ ID NO 1180; 980pp; English.  
XX  
CC The invention relates to isolated nucleic acid molecules and their  
CC encoded secreted proteins. The nucleic acids and proteins are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. Antibodies to the proteins can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunosorbent assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischemia, angiogenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection, and many other  
CC disorders listed in the specification. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence represents a novel secreted protein of the invention.

Alignment Scores:  
Pred. No.: 4,47e-80 Length: 162  
Score: 727.00 Matches: 137  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 36.12% Indels: 0  
DB: 22 Gaps: 0

US-09-763-902b-15 (1-1146) x AAU16227 (1-162)

OY 170 GATTTAAACCTTAGATGATGAGCATGTAGAGCTGCTGAGAGTTTGTCAATATTAC 229  
DB 26 AspphelysthrTyrtalaspnlnalacysargalalaglcluhelvalanlietyr 45  
OY 230 TATGACACATGATTAAGAGAGCGGACATPACCGAGCTGTATGTGACAAGCCACC 289  
DB 46 TyrgluthrmetaspysarfgarfgalaleuthrArgluetyrleuaspplsalyalatr 65  
OY 290 TTAATATGAGATGAAATGCTGTTTCAGGCGCTGAGCGCTTAATATTTTGTGACA 349  
DB 66 LeuiletrpaspnclyasnalaValserglyleuaspAlaleuasnnsnpherpaspntr 85  
OY 350 TTGCCTTAGTAGAGTCCAGGCTCAATATGTAGATTGCCAACAGTTATGAGACAGA 409

DB 86 LeuprosersercluhneglnalasnmetleuaspCysglmprovAlhISgluglnala 105  
OY 410 ACTCAGTCCCAAACTACAGTCTGTGTGACCACTGGAACGTGAGTTGATGGAAC 469  
DB 106 ThglInserglInthrThrValleuValThrSerGlyThrVallyspheaspClyasn 125  
OY 470 AAACACATTTCTTCACACGAACTCCGTGACGCTCAGTCCAGTCCCAAGTAAT 529  
DB 126 LysglInhISpheaspnslasnphleuThrAlaglInserThrProsnasntr 145  
OY 530 GTGTGAGATTCAGATGATGCTTCCTTTCAGATTGCTGCTAGTACT 580  
DB 146 ValtrpLyslelaleaserpsCyspneArpnehlaspTrpserSer 162  
RESULT 5  
AAU16228  
ID AAU16228 standard; Protein; 164 AA.  
XX  
AC AAU16228:  
XX  
DE 07-NOV-2001 (first entry)  
XX  
DT Human novel secreted protein, Seq ID 1181.  
XX  
XX Human; immunosuppressive; antiarthritic; antirheumatic;  
KW cytostatic; cardiant; vasotropic; cerebroprotective; nootropic;  
KW neuroprotective; antibacterial; virucide; fungicide; ophthalmological;  
KW vulnary; secreted protein; rheumatoid arthritis;  
KW hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
KW cerebrovascular disorder; cerebral ischemia; angiogenesis;  
KW nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
KW corneal infection; wound healing; epithelial cell proliferation;  
KW skin aging; food additive; preservative; antiproliferative.  
XX  
OS Homo sapiens.  
XX  
PN WO200155322-A2.  
XX  
PD 02-AUG-2001.  
XX  
PF 17-JAN-2001; 2001MO-US01341.  
XX  
PR 31-JAN-2000; 2000US-0179065.  
PR 04-FEB-2000; 2000US-0180628.  
PR 24-FEB-2000; 2000US-0184664.  
PR 02-MAR-2000; 2000US-0186350.  
PR 16-MAR-2000; 2000US-0189874.  
PR 17-MAR-2000; 2000US-0190076.  
PR 18-APR-2000; 2000US-0198123.  
PR 19-MAY-2000; 2000US-0205515.  
PR 07-JUN-2000; 2000US-0209467.  
PR 28-JUN-2000; 2000US-0214886.  
PR 30-JUN-2000; 2000US-0215135.  
PR 07-JUL-2000; 2000US-0216647.  
PR 07-JUL-2000; 2000US-0216880.  
PR 11-JUL-2000; 2000US-0217487.  
PR 11-JUL-2000; 2000US-0217496.  
PR 14-JUL-2000; 2000US-0218290.  
PR 26-JUL-2000; 2000US-0220963.  
PR 26-JUL-2000; 2000US-0220964.  
PR 14-AUG-2000; 2000US-0224518.  
PR 14-AUG-2000; 2000US-0224519.  
PR 14-AUG-2000; 2000US-0225213.  
PR 14-AUG-2000; 2000US-0225214.  
PR 14-AUG-2000; 2000US-0225266.  
PR 14-AUG-2000; 2000US-0225267.  
PR 14-AUG-2000; 2000US-0225268.  
PR 14-AUG-2000; 2000US-0225270.  
PR 14-AUG-2000; 2000US-0225447.  
PR 14-AUG-2000; 2000US-0225457.  
PR 14-AUG-2000; 2000US-0225757.  
PR 14-AUG-2000; 2000US-0225758.  
PR 14-AUG-2000; 2000US-0225759.

	PR	18-AUG-2000;	2000US-0226279.
	PR	22-AUG-2000;	2000US-0226681.
	PR	22-AUG-2000;	2000US-0226868.
	PR	22-AUG-2000;	2000US-0227182.
	PR	23-AUG-2000;	2000US-0227009.
	PR	30-AUG-2000;	2000US-0228924.
	PR	01-SEP-2000;	2000US-0229287.
	PR	01-SEP-2000;	2000US-0229343.
	PR	01-SEP-2000;	2000US-0229344.
	PR	01-SEP-2000;	2000US-0229345.
	PR	05-SEP-2000;	2000US-0229509.
	PR	05-SEP-2000;	2000US-0229513.
	PR	06-SEP-2000;	2000US-0230437.
	PR	06-SEP-2000;	2000US-0230438.
	PR	08-SEP-2000;	2000US-0231242.
	PR	08-SEP-2000;	2000US-0231243.
	PR	08-SEP-2000;	2000US-0231244.
	PR	08-SEP-2000;	2000US-0231413.
	PR	08-SEP-2000;	2000US-0231414.
	PR	08-SEP-2000;	2000US-0232080.
	PR	12-SEP-2000;	2000US-0232081.
	PR	12-SEP-2000;	2000US-0231968.
	PR	14-SEP-2000;	2000US-0232397.
	PR	14-SEP-2000;	2000US-0232398.
	PR	14-SEP-2000;	2000US-0232399.
	PR	14-SEP-2000;	2000US-0232400.
	PR	14-SEP-2000;	2000US-0232401.
	PR	14-SEP-2000;	2000US-0233063.
	PR	14-SEP-2000;	2000US-0233064.
	PR	14-SEP-2000;	2000US-0233065.
	PR	21-SEP-2000;	2000US-0234223.
	PR	21-SEP-2000;	2000US-0234274.
	PR	25-SEP-2000;	2000US-0234997.
	PR	25-SEP-2000;	2000US-0234998.
	PR	26-SEP-2000;	2000US-0235484.
	PR	27-SEP-2000;	2000US-0235834.
	PR	27-SEP-2000;	2000US-0235836.
	PR	29-SEP-2000;	2000US-0236327.
	PR	29-SEP-2000;	2000US-0236367.
	PR	29-SEP-2000;	2000US-0236368.
	PR	29-SEP-2000;	2000US-0236369.
	PR	29-SEP-2000;	2000US-0236370.
	PR	02-OCT-2000;	2000US-0236802.
	PR	02-OCT-2000;	2000US-0237037.
	PR	02-OCT-2000;	2000US-0237038.
	PR	02-OCT-2000;	2000US-0237039.
	PR	02-OCT-2000;	2000US-0237040.
	PR	13-OCT-2000;	2000US-0239935.
	PR	13-OCT-2000;	2000US-0239937.
	PR	20-OCT-2000;	2000US-0240960.
	PR	20-OCT-2000;	2000US-0241221.
	PR	20-OCT-2000;	2000US-0241785.
	PR	20-OCT-2000;	2000US-0241786.
	PR	20-OCT-2000;	2000US-0241787.
	PR	20-OCT-2000;	2000US-0241808.
	PR	20-OCT-2000;	2000US-0241809.
	PR	20-OCT-2000;	2000US-0241826.
	PR	01-NOV-2000;	2000US-0244617.
	PR	08-NOV-2000;	2000US-0246474.
	PR	08-NOV-2000;	2000US-0246475.
	PR	08-NOV-2000;	2000US-0246476.
	PR	08-NOV-2000;	2000US-0246477.
	PR	08-NOV-2000;	2000US-0246478.
	PR	08-NOV-2000;	2000US-0246523.
	PR	08-NOV-2000;	2000US-0246524.
	PR	08-NOV-2000;	2000US-0246525.
	PR	08-NOV-2000;	2000US-0246526.
	PR	08-NOV-2000;	2000US-0246527.
	PR	08-NOV-2000;	2000US-0246528.
	PR	08-NOV-2000;	2000US-0246532.
	PR	08-NOV-2000;	2000US-0246609.
	PR	08-NOV-2000;	2000US-0246610.
	PR	08-NOV-2000;	2000US-0246611.
	PR	08-NOV-2000;	2000US-0246611.
	PR	08-NOV-2000;	2000US-0246613.
	PR	17-NOV-2000;	2000US-0249207.
	PR	17-NOV-2000;	2000US-0249208.
	PR	17-NOV-2000;	2000US-0249209.
	PR	17-NOV-2000;	2000US-0249210.
	PR	17-NOV-2000;	2000US-0249211.
	PR	17-NOV-2000;	2000US-0249212.
	PR	17-NOV-2000;	2000US-0249213.
	PR	17-NOV-2000;	2000US-0249214.
	PR	17-NOV-2000;	2000US-0249215.
	PR	17-NOV-2000;	2000US-0249216.
	PR	17-NOV-2000;	2000US-0249217.
	PR	17-NOV	

Pred. No.: 1,47e-62 Length: 164  
Score: 584.50 Matches: 106  
Percent Similarity: 90.85% Conservative: 23  
Best Local Similarity: 74.65% Mismatches: 12  
Query Match: 29.04% Indels: 1  
DB: 22 Gaps: 1

US-09-763-902b-15 (1-1146) x AAU16228 (1-164)

QY 152 CAGATGGCCACGCTCTGATTTAAACTATATAGATCGCATGTAGAGCTGCTGAG 211  
D 24 Glnuetaia---ServalaspheleysthryValaspGlnalacysargalalaGlu 42  
QY 212 GAGTTTTCATATTTTACTATGACATGATGATAAGAGAGCGGACCTACACGAGCTG 271  
D 43 GlupheValasnvalITrYrThrThmetaspIysargalargleuSeIargleu 62  
QY 272 TATCTGACACAGCGCCACTTAATATGAAATGAAATGCTGTTGAGGGCTGATGCCCTA 331  
D 63 TyrmctgIyThralatThrleuValtrpAsnGlyasnalaValserGlyGlnIuserleu 82  
QY 332 AATATTTTTCGACATGCTCTTCTAGTGTCCAGGTCATATGTGATTTGCCAA 391  
D 83 SerGluPhepGluMetleuProserSerGluPheGlnIleSerValalaspGlyGln 102  
QY 392 CCAGTTCATGACAGACACGACGACGACGACGACGACGACGACGACGACGACGACG 451  
D 103 ProvalHisaspGluValatThrProserGlnThrThrValleuValalIleCysGlySer 122  
QY 452 GTGAGTTTGTATGAAACAAACAAACATTTCTTCAACGAGACTTCTGCTGCTGAG 511  
D 123 ValysPheGluGlnGlyasnGlyGlnArgaspheasnGlnanPheleuThralGln 142  
QY 512 TCCACTCCCAACAACTACTGTGTGGAAGATTCGAAGTGTCTTCCGTTTCAGATTGG 571  
D 143 AlaserProserAsnThrValtrpIlysilalaserAspCysPheargPheGlnaspTrp 162  
QY 572 TCTAGT 577  
D 163 Alaser 164

RESULT 6  
AA82318  
ID AA82318 standard; Protein: 140 AA.

XX AAY82318;  
AC  
XX 19-JUN-2000 (first entry)  
DT  
XX  
DE Human protein transport molecule (PTAM). SEQ ID NO:2.  
XX  
KW Human; protein transport molecule; PTAM; diagnosis; cytostatic;  
KW antihistaminic; antidiabetic; immunosuppressant; antileukosclerotic;  
KW antiallergic; antidiabetic; antileukemic; antirheumatic; osteoporotic;  
KW dermatological; antineoplastic; antiproliferative; hepatotropic; antigout;  
KW antineoplastic; antihistaminic; antiproliferative; protein transport regulator; cancer;  
KW immune disorder; cell proliferative disorder; secretory disorder;  
KW urticaria; allergy; abnormal vesicle trafficking; asthma;  
KW autoimmune haemolytic anaemia.  
XX  
XX Homo sapiens.  
OS  
XX  
XX WO200012703-A2.  
PN  
XX  
XX 09-MAR-2000.  
PD  
XX  
XX 26-AUG-1999; 99WO-US19616.  
PE  
XX  
XX 27-AUG-1998; 98US-0098206.  
PR  
XX  
XX (INCY-) INCYTE PHARM INC.  
PA  
XX  
XX Tang YT, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ.

PI Gorgone GA, Baughn MR, Patterson C;  
XX  
XX WPI: 2000-256642/22.  
DR  
DR N-PSDB: AAA08036.  
XX  
XX New human protein transport-associated polypeptide and polynucleotide  
PT useful for diagnosis, prevention and treatment of cell proliferative  
PT and secretory disorders such as leukemia, cystic fibrosis  
XX  
XX Claim 1: Page 61-62; 75pp; English.  
XX  
XX AAA08035 to AAA08042 encode the human protein transport-associated  
CC molecules (PTAMs) given in AA82317 to AA82324. The PTAMs have  
CC cytosolic, antihistaminic, antidiabetic, immunosuppressant,  
CC antileukosclerotic, antiallergic, antidiabetic, antileukemic,  
CC antineoplastic, osteoporotic, dermatological, antineoplastic, antiproliferative,  
CC hepatotropic, antigout, antineoplastic and antihistaminic activities, and  
CC regulate protein transport. PTAM proteins and antagonists are useful for  
CC preventing or treating a disorder associated with decreased or increased  
CC expression or activity of PTAM. PTAM polynucleotides are useful for  
CC diagnosing conditions associated with PTAM, comprising detecting PTAM by  
CC forming a hybridisation complex, preferably after PCR amplifying the  
CC biological sample. Diseases prevented, treated or diagnosed include cell  
CC proliferative disorders such as cancers, immune disorders, secretory  
CC disorders and other conditions associated with abnormal vesicle  
CC trafficking, such as allergies, asthma, urticaria and autoimmune  
CC haemolytic anaemia. Anti-PTAM antibodies may be used as antagonists, as  
CC a targeting or delivery mechanism for bringing pharmaceutical agents  
CC into cells or tissues expressing PTAM and for diagnosis of PTAM-related  
CC disorders. PTAM, its catalytic or immunogenic fragments are useful for  
CC drug screening using libraries of compounds. PTAM polynucleotides are  
CC useful for generating hybridisation probes useful in mapping the  
CC naturally occurring genomic sequences.  
XX  
XX Sequence 140 AA:  
SQ

Alignment Scores:  
Pred. No.: 1,58e-62 Length: 140  
Score: 584.00 Matches: 104  
Percent Similarity: 91.30% Conservative: 22  
Best Local Similarity: 75.36% Mismatches: 12  
Query Match: 29.01% Indels: 0  
DB: 21 Gaps: 0

US-09-763-902b-15 (1-1146) x AA82318 (1-140)

QY 164 TCTGTGATTTTAAACTATATAGATCGCATGTAGAGCTGAGAGTTTGTCAAT 223  
D 3 ServalaspheleysthryValaspGlnalacysargalalaGlnGluPheValasn 22  
QY 224 ATTTACTGTGACATGATGATTAAGAAGAGCGGACCTAACAGGCTTATCTGACAAAG 283  
D 23 ValTrpTrpThrThmetaspIysargalargleuSeIargleuTrpMetGlyThr 42  
QY 284 GCCACCTTAATATGAAATGAAATGCTTTCAGGCGTGATGCCCTAAATATTTT 343  
D 43 AlatrleuValtrpAsnGlyasnalaValserGlyGlnIuserleuSeIleuPhe 62  
QY 344 GACACATTTGCTTCTAGTGTGATTCAGGCTCAATATGATGATGCAACAGTTGATGAG 403  
D 63 GlnuetaiaProserSerGluPheGlnIleSerValalaspCysGlnProvalHisasp 82  
QY 404 CAAGCACTGACGCTCCCAACTAGACGCTTCTGTGTGACAGTGAAGTGAAGTTTGTAT 463  
D 83 GluValatThrProserGlnThrThrValleuValalIleCysGlySerValIysPheGln 102  
QY 464 GGAACCAACACATTTTCTTCAACGACATTTCCGCTGATGCTGATCCACCTCCAC 523  
D 103 GlyasnGlyGlnArgaspheasnGlnanPheleuThralalalaSerProser 122  
QY 524 AATACTGTGTGAGATGACAGTATGATGCTTCCGTTTCAAGATGCTAGT 577  
D 123 AsnThrValtrpIlysilalaserAspCysPheargPheGlnaspTrpAlaser 140

[illegible]

OY		350 TTGCGCTTTAGACGTTCACGCTCATATGATTGATGCATCCCAAGTTCATGAGCAACA	409
Dd		64 LeuProSerSerAsnHisGlnLeuAsnThrLeuAspAlaGlnProIleValAspGlnAla	83
OY		410 ACTCAGTCCCAACTACAGTTCCTTGTTGTGACCAGTAGCAAGCTGTGAATGGTAATGGAAC	469
Dd		84 ValSerAsnGlnLeuAlaArgLeuIleMetAlaSerGlySerValLysPheAlaAspGln	103
OY		470 AAACAACATTTCTTTCACCAACACTCCGCTGACGTGCTCAGTCCACTCCCAACATACT	529
OY		104 GlnLeuArgLysPheGlnGlnInrPhrIleValAlaThrAlaGlu-----Asnasp	119
Dd		530 GTGTGAGATGTCAGATGATGATGCTGCCCTTTTCAGAT	568
		120 LysTrpLysValValSerAspCysTyraArgMetGlnGlu	132
 RESULT 8 ABB67404 ID ABB67404 standard; Protein: 688 AA.			
XX	ABB67404:		
XX			
XX	26-MAR-2002 (first entry)		
XX			
DE	Drosophila melanogaster polypeptide seq ID NO 29004.		
KM	Drosophila: developmental biology; cell signalling; insecticide;		
XX	pharmaceutical.		
OS	Drosophila melanogaster.		
PN	MO2001J1042-A2.		
PD	27-SEP-2001.		
PF	23-MAR-2001; 2001WO-US09231.		
PR	23-MAR-2000; 2000US-191637P.		
XX	11-JUL-2000; 2000US-0614450.		
PA	(PEKE ) PE CORP NY.		
PI	Venter JC, Adams M, Li PWD, Myers EW;		
DR	WPI: 2001-656860/75.		
XX	N-PSDB; ABL11507.		
PT	New isolated nucleic acid detection reagent for detecting 1000 or more		
PT	genes from Drosophila and for elucidating cell signalling and cell-cell		
PT	interactions -		
PS	Disclosure: SEQ ID NO 29004; 21pp + Sequence Listing; English.		
XX			
CC	The invention relates to an isolated nucleic acid detection reagent		
CC	capable of detecting 1000 or more genes from Drosophila. The invention is		
CC	useful in developmental biology and in elucidating cell signalling and		
CC	cell-cell interactions in higher eukaryotes for the development of		
CC	insecticides, therapeutics and pharmaceutical drugs. The invention		
CC	discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA		
CC	sequences (ABB57737-ABB72072).		
CC	(ABB57737-ABB72072).		
CC	The sequence data for this patent did not form part of the printed		
CC	specification, but was obtained in electronic format directly from WIPO		
CC	at ftp://ftp.int/pub/published/PCT-sequences.		
XX	Sequence 688 AA;		
 Alignment Scores:			
Pred. No.:	0.000184	Length:	688
Score:	114.00	Matches:	43
Percent Similarity:	39.868	Conservative:	16
Best Local Similarity:	29.05%	Mismatches:	63

	Query Match:	5	66%	IndeIs:	26
	Dn:	22		Gaps:	6
	US-09-763-902B-15 (1-1146) x ABB67404 (1-688)				
OY	140 GCTCAAGATGCGCCAGATGGGCCACGTCGTGCAGATTAAACTTANTGTAGATCAGCATGT	199			
	.....				
Dd	3 AlAtThrGlInSeRgLnPrObSerProGlnSer-----	13			
OY	200 AGACSTGCTGAAGACTTTGTCAATAATTACTATGAGACAATGGATTAAGAAGACGGGCA	259			
Dd	14 ---ValGIYAyguIurPheValarGlnTYTTrILeUdaSnLysAlARoAsnHls	32			
OY	260 CTAACSCAGGCTGTATGAGACAGGCCACCTTAATA--TGAAATGGAATGCTGTTC	316			
	:				
Dd	33 LeuNlsArgrPheTyAasnlHisAnsSerSeryTLleNIsgLYubSerLysLeuVal	52			
OY	317 GGCGCGGATGCCCTTAATAATTTTTTGTACACATTCGCTTCAGTAGTCCAGGTCAAT	376			
	:				
Dd	53 GlYlNaIrYguLIuLLIHlsAnArglllegln-----GlnLeuAsn	65			
OY	377 ATGTAGATTGC-----CAACCAGTTCATGACAGCAAGCAGTCAGTCCCAAAGT	424			
Dd	66 PheAsnaAPCyshnlAlalylslleserGlnValAspAlaglnAlThrLeugLyasnGLY	85			
OY	425 ACAGTGTCTGTGTGACACAGTCGAACGTGTGAAGTTGATGAGAACAAACACATTTCTTC	484			
	:				
Dd	86 ValValValIGlnValThr---GlYluLeuSerAsnAspGlyGlnPrometargArprhe	104			
OY	485 AACSCGAAGTCCCTGCTGACATGCTGATCCACTCCCAACAANACTGTGTGGAAGATTGCA	544			
Dd	105 ThrGlnThrPheValLeuAlaLaGlnSerProLysLys-----TyrTYrValHls	121			
OY	545 AGTGATGCTTCCTCGTTTCCAAGAT	568			
	:::             ::				
Dd	122 AsnAspIlePheArgTYrGlnasp	129			
	RESULT 9				
ID	ABB66052				
AC	ABB66052 standard; Protein; 690 AA.				
XX	ABB66052;				
DT	26-MAR-2002 (first entry)				
DE	Drosophila melanogaster polypeptide SHQ ID NO 24948.				
KW	Drosophila; developmental biology; cell signalling; insecticide;				
RK	pharmaceutical.				
OS	Drosophila melanogaster.				
PX	WO200171042-A2.				
PD	27-SEP-2001.				
PF	23-MAR-2001; 2001WO-US09231.				
PR	23-MAR-2000; 2000US-191637P.				
PT	11-JUL-2000; 2000US-0614150.				
FA	(PEKE ) PE CORP NY.				
PI	Venter JC, Adams M, Li PWD, Myers EW;				
DR	WPI; 2001-656860/75.				
N-PSDB	ABLI0155.				
XX	New isolated nucleic acid detection reagent for detecting 1000 or more				
XX	genes from Drosophila and for elucidating cell signalling and cell-cell				
XX	interactions -				
XX	Disclosure: SEQ ID NO 24948; 21pp + Sequence Listing; English..				

[illegible]



PN EPI033405-A2.  
 XX 06-SEP-2000.  
 XX 25-FEB-2000; 2000EP-0301439.  
 XX 25-FEB-1999; 9905-0121825.  
 PR 05-MAR-1999; 9905-0123180.  
 PR 09-MAR-1999; 9905-0123548.  
 PR 23-MAR-1999; 9905-0125788.  
 PR 25-MAR-1999; 9905-0126264.  
 PR 29-MAR-1999; 9905-0126785.  
 PR 01-APR-1999; 9905-0127462.  
 PR 06-APR-1999; 9905-0128234.  
 PR 08-APR-1999; 9905-0128714.  
 PR 16-APR-1999; 9905-0129845.  
 PR 19-APR-1999; 9905-0130077.  
 PR 21-APR-1999; 9905-0130449.  
 PR 23-APR-1999; 9905-0130510.  
 PR 23-APR-1999; 9905-0130891.  
 PR 28-APR-1999; 9905-0131449.  
 PR 30-APR-1999; 9905-0132048.  
 PR 30-APR-1999; 9905-0132407.  
 PR 04-MAY-1999; 9905-0132484.  
 PR 05-MAY-1999; 9905-0132485.  
 PR 06-MAY-1999; 9905-0132486.  
 PR 06-MAY-1999; 9905-0132487.  
 PR 07-MAY-1999; 9905-0132863.  
 PR 11-MAY-1999; 9905-0134256.  
 PR 14-MAY-1999; 9905-0134218.  
 PR 14-MAY-1999; 9905-0134219.  
 PR 14-MAY-1999; 9905-0134221.  
 PR 14-MAY-1999; 9905-0134370.  
 PR 18-MAY-1999; 9905-0134768.  
 PR 19-MAY-1999; 9905-0134941.  
 PR 20-MAY-1999; 9905-0135114.  
 PR 21-MAY-1999; 9905-0135353.  
 PR 24-MAY-1999; 9905-0135629.  
 PR 25-MAY-1999; 9905-0136021.  
 PR 27-MAY-1999; 9905-0136392.  
 PR 28-MAY-1999; 9905-0136782.  
 PR 01-JUN-1999; 9905-0137222.  
 PR 03-JUN-1999; 9905-0137528.  
 PR 04-JUN-1999; 9905-0137502.  
 PR 07-JUN-1999; 9905-0137724.  
 PR 08-JUN-1999; 9905-0138094.  
 PR 10-JUN-1999; 9905-0138540.  
 PR 10-JUN-1999; 9905-0138847.  
 PR 14-JUN-1999; 9905-0139119.  
 PR 16-JUN-1999; 9905-0139452.  
 PR 16-JUN-1999; 9905-0139453.  
 PR 17-JUN-1999; 9905-0139454.  
 PR 18-JUN-1999; 9905-0139455.  
 PR 18-JUN-1999; 9905-0139456.  
 PR 18-JUN-1999; 9905-0139457.  
 PR 18-JUN-1999; 9905-0139458.  
 PR 18-JUN-1999; 9905-0139459.  
 PR 18-JUN-1999; 9905-0139460.  
 PR 18-JUN-1999; 9905-0139461.  
 PR 18-JUN-1999; 9905-0139462.  
 PR 18-JUN-1999; 9905-0139463.  
 PR 18-JUN-1999; 9905-0139763.  
 PR 18-JUN-1999; 9905-0139763.  
 PR 21-JUN-1999; 9905-0139817.  
 PR 22-JUN-1999; 9905-0139819.  
 PR 23-JUN-1999; 9905-0140353.  
 PR 23-JUN-1999; 9905-0140354.  
 PR 24-JUN-1999; 9905-0140695.  
 PR 26-JUN-1999; 9905-0140823.  
 PR 29-JUN-1999; 9905-0140991.  
 PR 30-JUN-1999; 9905-0141287.  
 PR 01-JUL-1999; 9905-0141842.

PR 01-JUL-1999; 9905-0142154.  
 PR 02-JUL-1999; 9905-0142055.  
 PR 06-JUL-1999; 9905-0142390.  
 PR 08-JUL-1999; 9905-0142803.  
 PR 09-JUL-1999; 9905-0142920.  
 PR 12-JUL-1999; 9905-0142977.  
 PR 13-JUL-1999; 9905-0143542.  
 PR 14-JUL-1999; 9905-0143624.  
 PR 15-JUL-1999; 9905-0144005.  
 PR 16-JUL-1999; 9905-0144085.  
 PR 16-JUL-1999; 9905-0144086.  
 PR 19-JUL-1999; 9905-0144335.  
 PR 19-JUL-1999; 9905-0144331.  
 PR 19-JUL-1999; 9905-0144332.  
 PR 19-JUL-1999; 9905-0144333.  
 PR 19-JUL-1999; 9905-0144334.  
 PR 19-JUL-1999; 9905-0144335.  
 PR 20-JUL-1999; 9905-0144352.  
 PR 20-JUL-1999; 9905-0144632.  
 PR 20-JUL-1999; 9905-0144884.  
 PR 21-JUL-1999; 9905-0144814.  
 PR 21-JUL-1999; 9905-0145086.  
 PR 22-JUL-1999; 9905-0145087.  
 PR 22-JUL-1999; 9905-0145085.  
 PR 22-JUL-1999; 9905-0145089.  
 PR 22-JUL-1999; 9905-0145192.  
 PR 23-JUL-1999; 9905-0145145.  
 PR 23-JUL-1999; 9905-0145218.  
 PR 26-JUL-1999; 9905-0145224.  
 PR 27-JUL-1999; 9905-0145276.  
 PR 27-JUL-1999; 9905-0145591.  
 PR 27-JUL-1999; 9905-0145591.  
 PR 28-JUL-1999; 9905-0145951.  
 PR 02-AUG-1999; 9905-0146386.  
 PR 02-AUG-1999; 9905-0146388.  
 PR 02-AUG-1999; 9905-0146389.  
 PR 03-AUG-1999; 9905-0147038.  
 PR 04-AUG-1999; 9905-0147204.  
 PR 04-AUG-1999; 9905-0147302.  
 PR 05-AUG-1999; 9905-0147192.  
 PR 05-AUG-1999; 9905-0147260.  
 PR 06-AUG-1999; 9905-0147303.  
 PR 06-AUG-1999; 9905-0147416.  
 PR 09-AUG-1999; 9905-0147483.  
 PR 09-AUG-1999; 9905-0147935.  
 PR 10-AUG-1999; 9905-0148171.  
 PR 11-AUG-1999; 9905-0148319.  
 PR 12-AUG-1999; 9905-0148341.  
 PR 13-AUG-1999; 9905-0148565.  
 PR 13-AUG-1999; 9905-0148684.  
 PR 16-AUG-1999; 9905-0149388.  
 PR 17-AUG-1999; 9905-0149175.  
 PR 18-AUG-1999; 9905-0149426.  
 PR 20-AUG-1999; 9905-0149722.  
 PR 20-AUG-1999; 9905-0149723.  
 PR 20-AUG-1999; 9905-0149929.  
 PR 23-AUG-1999; 9905-0149902.  
 PR 23-AUG-1999; 9905-0149930.  
 PR 25-AUG-1999; 9905-0150566.  
 PR 26-AUG-1999; 9905-0150884.  
 PR 27-AUG-1999; 9905-0151065.  
 PR 27-AUG-1999; 9905-0151066.  
 PR 27-AUG-1999; 9905-0151080.  
 PR 30-AUG-1999; 9905-0151303.  
 PR 31-AUG-1999; 9905-0151930.  
 PR 01-SEP-1999; 9905-0151938.  
 PR 07-SEP-1999; 9905-0152363.  
 PR 10-SEP-1999; 9905-0153070.  
 PR 13-SEP-1999; 9905-0153758.  
 PR 15-SEP-1999; 9905-0154018.  
 PR 16-SEP-1999; 9905-0154039.

PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161982.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Alignment Scores:  
 Pred. No.: 0.0001  
 Score: 113.50  
 Percent Similarity: 46.02%  
 Best Local Similarity: 27.43%  
 Query Match: 5.64%  
 DB: 21  
 Gaps: 3

US-09-763-902b-15 (1-1146) x MAG12073 (1-125)  
 QY 206 GCTGAGAGAGTTGTCATATTTACTATGAGACAATGATTAAGAAGAGCGGACTAAC 265  
 Db 11 AAlaserlarphrevalasnhistyltlythistyleurpheaspranasparySerSerleuser 30  
 QY 266 AGGCTGATCTGACAGGCGCCCTTAATATGATGAAATGCTGTTCAGGGCTGAT 325  
 Db 31 SerleuserlyasprothInserleuserhphreglucgluInthlletylglyalasp 50  
 QY 326 GCCCTAAATATTTTTTTGACACATGCGCTTCTAGTGAAGTTCACG-----GTCAATATG 379  
 Db 51 Asnleuseranlylsleuylsglnleuprophesapngcysnhislsleuileserthr 70  
 QY 380 TTAGATGCCACAGAGTTGATGACAGACAGCTACGCCAACAGTCTTTGTG 439  
 Db 71 Valasberginpro-----SerSeretlaaglyglycylglylYlleuvalphe 88  
 QY 440 ACCAGTGAAGTGAAGTTGATGAGAAACAACAATTC--TTCACAGAGACTTC 496  
 Db 89 ValserglyserIleaglInleuInlglylulaspnhlsproleuargpheserGlnvaltyr 108  
 QY 497 CTGCTAGCTGTCTAGTGCACATCCCAACAATCTGTGTG 535

Db 109 LeuSerIleCysAsnArgPheGluGlyPheIleTrp 121  
 RESULT 11  
 ID MAG05145 standard; Protein; 153 AA.  
 AC MAG05145;  
 DT 17-OCT-2000 (first entry)  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 1441.  
 KW Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 OS Arabidopsis thaliana.  
 PN EP1033405-A2.  
 PD 06-SEP-2000.  
 XX 25-FEB-2000; 2000EP-0301439.  
 PF 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0122180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 30-APR-1999; 99US-0132407.  
 PR 04-MAY-1999; 99US-0132484.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132863.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-0134219.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 14-MAY-1999; 99US-0134370.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134941.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135621.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137528.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138094.  
 PR 10-JUN-1999; 99US-0138540.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.

PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 21-JUN-1999; 99US-0139753.  
PR 22-JUN-1999; 99US-0139889.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142970.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144844.  
PR 21-JUL-1999; 99US-0144844.  
PR 21-JUL-1999; 99US-0145086.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145218.  
PR 26-JUL-1999; 99US-0145224.  
PR 27-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148564.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.

PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 23-AUG-1999; 99US-0150366.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.: 0.000109 Length: 153  
Score: 113.50 Matches: 36  
Percent Similarity: 45.32% Conserved: 27  
Best Local Similarity: 25.908 Mismatches: 65  
Query Match: 5.64% Indels: 11  
DB: 21 Gaps: 5

US-09-763-902b-15 (1-1146) x ANG05145 (1-153)

OY 158 GCCACGCTCTGATTTAACTTATGTAGATCAGCATGTAGAGCTGCTGAGGACTTT 217  
DB 13 SerAlaSerLeuSerLeuLysMetSerGlnMetAspProAspAlaValSerLysAlaPhe 32

PR	18-MAY-1999	9905-0134768
PR	19-MAY-1999	9905-01354941
PR	20-MAY-1999	9905-01351244
PR	21-MAY-1999	9905-01355529
PR	24-MAY-1999	9905-01356529
PR	25-MAY-1999	9905-01360621
PR	27-MAY-1999	9905-01363692
PR	28-MAY-1999	9905-01366782
PR	01-JUN-1999	9905-01373222
PR	03-JUN-1999	9905-01375288
PR	04-JUN-1999	9905-01375602
PR	07-JUN-1999	9905-0137724
PR	08-JUN-1999	9905-0138094
PR	10-JUN-1999	9905-0138340
PR	10-JUN-1999	9905-01388477
PR	14-JUN-1999	9905-01391119
PR	16-JUN-1999	9905-01394512
PR	16-JUN-1999	9905-01394553
PR	17-JUN-1999	9905-01394942
PR	18-JUN-1999	9905-01394954
PR	18-JUN-1999	9905-01394660
PR	18-JUN-1999	9905-01394661
PR	18-JUN-1999	9905-01394662
PR	18-JUN-1999	9905-01394663
PR	18-JUN-1999	9905-01394664
PR	18-JUN-1999	9905-01394665
PR	18-JUN-1999	9905-01394666
PR	18-JUN-1999	9905-01394667
PR	18-JUN-1999	9905-01394668
PR	18-JUN-1999	9905-01394669
PR	18-JUN-1999	9905-01394670
PR	18-JUN-1999	9905-01394671
PR	18-JUN-1999	9905-01394672
PR	18-JUN-1999	9905-01394673
PR	18-JUN-1999	9905-01394674
PR	18-JUN-1999	9905-01394675
PR	18-JUN-1999	9905-01394676
PR	18-JUN-1999	9905-01394677
PR	18-JUN-1999	9905-01394678
PR	18-JUN-1999	9905-01394679
PR	18-JUN-1999	9905-01394680
PR	18-JUN-1999	9905-01394681
PR	18-JUN-1999	9905-01394682
PR	18-JUN-1999	9905-01394683
PR	18-JUN-1999	9905-01394684
PR	18-JUN-1999	9905-01394685
PR	18-JUN-1999	9905-01394686
PR	18-JUN-1999	9905-01394687
PR	18-JUN-1999	9905-01394688
PR	18-JUN-1999	9905-01394689
PR	18-JUN-1999	9905-01394690
PR	18-JUN-1999	9905-01394691
PR	18-JUN-1999	9905-01394692
PR	18-JUN-1999	9905-01394693
PR	18-JUN-1999	9905-01394694
PR	18-JUN-1999	9905-01394695
PR	18-JUN-1999	9905-01394696
PR	18-JUN-1999	9905-01394697
PR	18-JUN-1999	9905-01394698
PR	18-JUN-1999	9905-01394699
PR	18-JUN-1999	9905-01394700
PR	18-JUN-1999	9905-01394701
PR	18-JUN-1999	9905-01394702
PR	18-JUN-1999	9905-01394703
PR	18-JUN-1999	9905-01394704
PR	18-JUN-1999	9905-01394705
PR	18-JUN-1999	9905-01394706
PR	18-JUN-1999	9905-01394707
PR	18-JUN-1999	9905-01394708
PR	18-JUN-1999	9905-01394709
PR	18-JUN-1999	9905-01394710
PR	18-JUN-1999	9905-01394711
PR	18-JUN-1999	9905-01394712
PR	18-JUN-1999	9905-01394713
PR	18-JUN-1999	9905-01394714
PR	18-JUN-1999	9905-01394715
PR	18-JUN-1999	9905-01394716
PR	18-JUN-1999	9905-01394717
PR	18-JUN-1999	9905-01394718
PR	18-JUN-1999	9905-01394719
PR	18-JUN-1999	9905-01394720
PR	18-JUN-1999	9905-01394721
PR	18-JUN-1999	9905-01394722
PR	18-JUN-1999	9905-01394723
PR	18-JUN-1999	9905-01394724
PR	18-JUN-1999	9905-01394725
PR	18-JUN-1999	9905-01394726
PR	18-JUN-1999	9905-01394727
PR	18-JUN-1999	9905-01394728
PR	18-JUN-1999	9905-01394729
PR	18-JUN-1999	9905-01394730
PR	18-JUN-1999	9905-01394731
PR	18-JUN-1999	9905-01394732
PR	18-JUN-1999	9905-01394733
PR	18-JUN-1999	9905-01394734
PR	18-JUN-1999	9905-01394735
PR	18-JUN-1999	9905-01394736
PR	18-JUN-1999	



PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130049.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132487.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 17-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142820.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.

PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147483.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.





PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145919.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.

PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152363.  
PR 10-SEP-1999; 99US-0153070.  
PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.: 0.000136  
Score: 112.50  
Percent Similarity: 47.97%  
Best Local Similarity: 26.83%  
Query Match: 5.59%  
DB: 21  
Gaps: 5

US-09-763-902b-15 (1-1146) x AAG05147 (1-131)

QY 206 GCTGAGAGCTTGTCAATATTTACTATGAGCAATGATGATAAAGAGCGGCGCTAAC 265  
DB 7 SerLysAlaPheValGluHisTyrSerThrPheAspThrAsnArgValGlyLeuAla 26  
QY 266 AGCGTGTACTCTGAGCAAGCCACCTTAATATGATGAAATGCTGTTTCAGCGCTGAT 325  
DB 27 GlyLeuTyrGlnGluAlaSerMetLeuThrPheGluGlnLysIleGlnGlyAlaGln 46  
QY 326 GCCCTAATATATTTTTCACATTCGCT-----TCAGTACGCTTCAGGCTCATATG 379  
DB 47 SerIleValAlaLysLeuThrSerLeuProPheGlnGlnCysLysHisIleSerThr 66

OY 380 TTGATGGCCACCACTTCATGACGACCACTCACTCCCAACTGATCTTGTGTG 439  
Db 67 ValAspCysGlnPro-----SerGlyProAlaSerGlyMetLeuValPhe 81  
OY 440 ACCAGTGGACACTGTGATGATGGAACAACAACAACTTC--TTCAACGAGAACTTC 496  
Db 82 ValSerGlyAsnLeuGlnLeuAlaGlyGluValSHisAlaLeuValPheSerGlnMetPhe 101  
OY 497 CTCCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGACGCTGAC 556  
Db 102 HisLeu-----MetProHisProGlnGlySerPheTyr--ValPheAsnAspIlePhe 118  
OY 557 CGTTTCAA 565  
Db 119 SerTrpArg 121  
RESULT 15  
AG05146  
ID AG05146 standard; Protein; 134 AA.  
XX  
AC AG05146:  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1442.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KM hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EPI033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.

PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144332.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.

PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 06-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 09-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0148171.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148565.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160880.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161922.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Alignment Scores:  
 Pred. No.: 0.000137 Length: 134  
 Score: 112.50 Matches: 33  
 Percent Similarity: 47.978 Conservative: 26  
 Best Local Similarity: 26.838 Mismatches: 53  
 Query Match: 5.598 Indels: 11  
 DB: 21 Gaps: 5

US-09-763-902b-15 (1-1146) x AAG05146 (1-134)

QY 206 GCTGAGAGCTTGTCTAATATTACTATGAGACAATGGATMAAAGAGAGCGGCACTAAC 265  
 : : : : :  
 Db 10 SerLysAlaPheValGlnHisTyrTyrSerThrPheAspThrAsnArgValGlyLeuAla 29  
 : : : : :  
 QY 266 AGGCTGTATCTGGACAGGCGCACTTAATATGGAATGGAATGCTGTTTCCAGGCTGAT 325  
 : : : : :  
 Db 30 GlyLeuTyrGlnGlnHisSerMetLeuThrPheGlnGlyGlnHisGlnGlyValGln 49  
 : : : : :  
 QY 326 GCCCTAATAATTTTGTGACATGCTCTTCTAGTGAGTTCAGGTCATATG 379  
 : : : : :  
 Db 50 SerIleValAlaLysLeuThrSerLeuProPheGlnGlnCysLysHisSerThr 69  
 : : : : :  
 QY 380 TTAGATTGCCAACCACTTCATGAGCAAGCACTGCTCCAACTGCTGTTGTTG 439  
 : : : : :  
 Db 70 ValAspCysGlnPro-----SerGlyProHisSerGlyMetLeuValPhe 84  
 : : : : :  
 QY 440 ACCAGTGAAGCTGTGATGATGGAACAACAACATTC--TTCACCAAGAACTTC 496  
 : : : : :  
 Db 85 ValSerGlyAsnLeuGlnLeuAlaGlyGlnLysHisAlaLeuLysSerGlnMetPhe 104  
 : : : : :  
 QY 497 CTCCTGACGCTCAGTCCAGTCCCAACATACGCTGTGGAAGTTCAGTATGCTTC 556  
 : : : : :  
 Db 105 HisLeu-----MetProThrProGlnGlySerPheTyr--ValPheAsnAspIlePhe 121  
 : : : : :  
 QY 557 CGTTTCAA 565  
 : : : : :  
 Db 122 SerTyrParg 124

Search completed: June 24, 2003, 19:31:07  
 Job time : 58.8062 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

# OM nucleic - nucleic search, using sw model

Run on: June 28, 2003, 01:01:07 ; Search time 138.193 Seconds  
(without alignments)  
8915.714 Million cell updates/sec

Title: US-09-763-902B-10  
Perfect score: 830  
Sequence: 1 tgactcttgacattcacctg.....ataattataataacacatg 830

Scoring table:  
IDENTITY\_NUC  
Gapop 10.0, Gapext 1.0

Searched: 1055720 seqs, 742224136 residues

Total number of hits satisfying chosen parameters: 2111440

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database : Published Applications NA.\*

```

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PC7_NEM_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEM_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEM_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PC7US_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEM_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_NEM_PUB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US10_NEM_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US60_NEM_PUB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	761	91.7	864	10 US-09-764-864-394	Sequence 394, App
2	318	38.3	318	9 US-09-796-692-4160	Sequence 4160, App
3	318	38.3	318	9 US-10-040-862-4160	Sequence 4160, App
4	317	38.2	318	10 US-09-815-343-1201	Sequence 1201, App
5	250	30.1	1387	10 US-09-764-864-393	Sequence 393, App
6	216.8	26.1	415	9 US-09-918-995-2901	Sequence 2901, App
7	171.8	20.7	417	9 US-09-918-995-6557	Sequence 6557, App
8	113	13.6	8918	10 US-09-764-864-1682	Sequence 1682, App
9	113	13.6	8919	10 US-09-764-864-1683	Sequence 1683, App
10	36.4	4.4	415	9 US-09-918-995-36691	Sequence 36691, A
11	36.4	4.4	88191	10 US-09-799-799-3	Sequence 3, Appl1
12	35.8	4.3	32191	10 US-09-764-877-3374	Sequence 3374, App
13	34.8	4.2	1141	9 US-10-008-119A-11	Sequence 11, Appl
14	34.8	4.2	1141	10 US-09-443-704-11	Sequence 11, Appl
15	34.8	4.2	1490	9 US-10-097-065-31	Sequence 31, Appl
16	34.8	4.2	172637	10 US-09-805-458A-3	Sequence 3, Appl1
17	34.2	4.1	555	9 US-10-123-155-72	Sequence 72, Appl
18	34	4.1	499	9 US-09-918-995-19696	Sequence 19696, A
19	33.6	4.0	1753	9 US-09-986-480-149	Sequence 149, App

20	33.6	4.0	5512	9 US-10-239-676-150	Sequence 150, App
21	33.6	4.0	684973	10 US-09-263-959-1	Sequence 1, Appl1
22	33.2	4.0	1294	9 US-09-938-842A-4254	Sequence 4254, App
23	33.2	4.0	1907	9 US-09-924-340-53	Sequence 53, Appl
24	33.2	4.0	1907	9 US-09-992-600A-53	Sequence 53, Appl
25	33.2	4.0	1907	9 US-10-000-489-53	Sequence 53, Appl
26	33.2	4.0	1907	9 US-10-000-986-53	Sequence 53, Appl
27	33.2	4.0	5881	9 US-10-172-086-81	Sequence 81, Appl
28	33.2	4.0	4852	10 US-09-864-761-15713	Sequence 15713, A
29	33	4.0	4059	12 US-10-042-417-47	Sequence 47, Appl
30	33	4.0	4221	9 US-10-125-237-70	Sequence 70, Appl
31	33	4.0	4221	9 US-10-105-891-70	Sequence 70, Appl
32	33	4.0	8764	9 US-09-764-891-10019	Sequence 10019, A
33	33	4.0	8911	12 US-10-044-090-17	Sequence 17, Appl
34	32.8	4.0	1503841	9 US-09-946-807-1	Sequence 1, Appl1
35	32.8	4.0	1503841	10 US-09-795-668-1	Sequence 1, Appl1
36	32.8	4.0	1503841	10 US-09-795-668-1	Sequence 1, Appl1
37	32.4	3.9	216	10 US-09-960-352-7902	Sequence 7902, App
38	32.2	3.9	239	10 US-09-919-580-208	Sequence 208, App
39	32.2	3.9	250	10 US-09-864-761-31695	Sequence 31695, A
40	32	3.9	457	9 US-09-918-995-11773	Sequence 11773, A
41	32	3.9	504	12 US-10-001-843-14	Sequence 14, Appl
42	32	3.9	517	10 US-09-919-580-874	Sequence 874, App
43	32	3.9	756	9 US-10-123-155-392	Sequence 392, App
44	32	3.9	1194	10 US-09-834-975-840	Sequence 840, App
45	32	3.9	3285	9 US-09-986-480-29	Sequence 29, Appl

## ALIGNMENTS

```

RESULT 1
US-09-764-864-394
Sequence 394, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT223
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 394
LENGTH: 864
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (1)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (7)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (9)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (15)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (31)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (837)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (853)
OTHER INFORMATION: n equals a,t,g, or c
NAME/KEY: SITE
LOCATION: (853)
OTHER INFORMATION: n equals a,t,g, or c

```

US-09-764-864-394

Query Match 91.7%; Score 761; DB 10; Length 864;  
Best Local Similarity 100.0%; Pred. No. 4.9e-240;  
Matches 761; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

70 GGAATACCCCTGGGAGAGCCCTCTTCATGAAACACAGATGGCATCTGTGATTCAA 129  
|||||  
35 GGAATACCCCTGGGAGAGCCCTCTTCATGAAACACAGATGGCATCTGTGATTCAA 94  
130 GACCTATGTGATAGAGGCTTCAGAGCTGAGAGAGTTTCTCATGCTCTACACAC 189  
|||||  
95 GACCTATGTGATAGAGGCTTCAGAGCTGAGAGAGTTTCTCATGCTCTACACAC 154  
190 CATGATTAAGGGGGGGGTTTGTCTCCCGCTTACATGAGGACAGACCCCTGTG 249  
155 CATGATTAAGGGGGGGGTTTGTCTCCCGCTTACATGAGGACAGACCCCTGTG 214  
250 GAATGCAATGCTGTTTCAGACAAAGATCCCTGAGTATGTTTGAATGTTCCCTC 309  
|||||  
215 GAATGCAATGCTGTTTCAGACAAAGATCCCTGAGTATGTTTGAATGTTCCCTC 274  
310 CAGGAGATCCCAATCAGCGGTAGACTGACACCTGTTTATGATGAGCCACACAG 369  
275 CAGGAGATCCCAATCAGCGGTAGACTGACACCTGTTTATGATGAGCCACACAG 334  
370 CAGGAGATCCCAATCAGCGGTAGACTGACACCTGTTTATGATGAGCCACACAG 429  
335 CAGGAGATCCCAATCAGCGGTAGACTGACACCTGTTTATGATGAGCCACACAG 394  
430 GGAATACCCCTGGGAGAGCCCTCTTCATGAAACACAGATGGCATCTGTGATTCAA 489  
|||||  
395 GGAATACCCCTGGGAGAGCCCTCTTCATGAAACACAGATGGCATCTGTGATTCAA 454  
490 GATGCAAGTGAAGCTGCTCCGCTTCAGAGAGTGGGAGAGTGGGAGAGTGGGAG 549  
455 GATGCAAGTGAAGCTGCTCCGCTTCAGAGAGTGGGAGAGTGGGAGAGTGGGAG 514  
550 TCTTGTCTCATTCAGCCCTAGCTCTGATGAGAAATCAACCTGCTGATGAGATG 609  
515 TCTTGTCTCATTCAGCCCTAGCTCTGATGAGAAATCAACCTGCTGATGAGATG 574  
610 GAGGACACAAATGATCTTGTGTTGGGAGAGACTGACAGACTCCACTGTGCGAGGT 669  
575 GAGGACACAAATGATCTTGTGTTGGGAGAGACTGACAGACTCCACTGTGCGAGGT 634  
670 TGAAGCTTTTGTGCTCAGAGTCTAGAGAGTCCCTTCTGATATATATCTTGTG 729  
635 TGAAGCTTTTGTGCTCAGAGTCTAGAGAGTCCCTTCTGATATATATCTTGTG 694  
730 CATAGTTCTCTTTCAGAGTAACTTCTTATTTTCTACTTGGCCAGTGAAGCTC 789  
695 CATAGTTCTCTTTCAGAGTAACTTCTTATTTTCTACTTGGCCAGTGAAGCTC 754  
790 TGAATTCGAAATTCGACAAATTAATTAATTAATTAATTAATTAATTAATTAAT 830  
755 TGAATTCGAAATTCGACAAATTAATTAATTAATTAATTAATTAATTAATTAAT 795

RESULT 2  
US-09-766-692-4160

Sequence 4160, Application US/09796692  
Publication No. US20020198362A1  
GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Algate, Paul A.  
APPLICANT: Mannion, Jane  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY  
FILE REFERENCE: 2077.001200  
CURRENT APPLICATION NUMBER: US/09/796,692  
CURRENT FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 60/186,126

PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: 60/206,201  
PRIOR FILING DATE: 2000-05-22, 950  
PRIOR APPLICATION NUMBER: 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: 60/223,378  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 9597  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4160  
LENGTH: 318  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-796-692-4160

Query Match 38.3%; Score 318; DB 9; Length 318;  
Best Local Similarity 100.0%; Pred. No. 2.8e-94;  
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

147 CTTGCAAGCTGCTGAGAGTTTGTCAATGCTCTACTACACCATGATGAGGGGGC 206  
1 CTTGCAAGCTGCTGAGAGTTTGTCAATGCTCTACTACACCATGATGAGGGGGC 60  
207 GTTGTGCTGCTCCGCTTATGATGAGGACACGACCTGCTGTGATGAGCAATGCTTT 266  
61 GTTGTGCTGCTCCGCTTATGATGAGGACACGACCTGCTGTGATGAGCAATGCTTT 120  
267 CAGGACAGAAATCCTGAGTGTGTTTGAATGTTGCTTCCAGGAGTCCAAATCA 326  
121 CAGGACAGAAATCCTGAGTGTGTTTGAATGTTGCTTCCAGGAGTCCAAATCA 180  
327 GCGTGTAGACTGCGACCTGTTTCATGATGAGGACACCAAGCCAGGAGTCCCTG 386  
181 GCGTGTAGACTGCGACCTGTTTCATGATGAGGACACCAAGCCAGGAGTCCCTG 240  
387 TTGTGATCTGTGATCAGTAACTTTGAGGGGAAACAACAAGGAGCTTCAACCAACT 446  
241 TTGTGATCTGTGATCAGTAACTTTGAGGGGAAACAACAAGGAGCTTCAACCAACT 300  
447 TCAATCTGACCGCCGAG 464  
301 TCAATCTGACCGCCGAG 318

RESULT 3  
US-10-040-862-4160

Sequence 4160, Application US/10040862  
Publication No. US20030078396A1  
GENERAL INFORMATION:  
APPLICANT: Gaiger, Alexander  
APPLICANT: Algate, Paul A.  
APPLICANT: Mannion, Jane  
APPLICANT: Retter, Marc  
TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Ther  
FILE REFERENCE: 014058-013520US

CURRENT APPLICATION NUMBER: US/10/040,862  
CURRENT FILING DATE: 2001-11-06  
PRIOR APPLICATION NUMBER: US 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: US 60/190,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: US 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: US 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: US 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: US 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: US 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: US 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: US 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: US 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: US 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: US 60/223,378  
PRIOR FILING DATE: 2000-08-07  
PRIOR APPLICATION NUMBER: US 09/796,692  
PRIOR FILING DATE: 2001-03-01  
NUMBER OF SEQ ID NOS: 10467  
SOFTWARE: FASTSEQ for Windows Version 3.0  
SEQ ID NO 4160  
LENGTH: 318  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-10-040-862-4160

Query Match  
Best Local Similarity 38.3%; Score 318; DB 9; Length 318;  
Matches 318; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 147 CCTGAGAGCTGTGAGAGAGTTTGTCAATGCTCTACTACACCAACGATGATAGCGGCGC 206  
DB 1 CCTGAGAGCTGTGAGAGAGTTTGTCAATGCTCTACTACACCAACGATGATAGCGGCGC 60  
QY 207 GTTGTCTGTCCCGCTGTACATGGGACAGCCACCCCTGTCTGAAATGCAATGCTGTT 266  
DB 61 GTTGTCTGTCCCGCTGTACATGGGACAGCCACCCCTGTCTGAAATGCAATGCTGTT 120  
QY 267 CAGACAGAATCTCTGATGATGTTTGTAAATGTTCCCTTCCAGCGAGTTCCAAATCA 326  
DB 121 CAGACAGAATCTCTGATGATGTTTGTAAATGTTCCCTTCCAGCGAGTTCCAAATCA 180  
QY 327 GCGGTGATGATGCGAGCGCTGTCTGATGATGAGGACACCAAGCAGAGGCTCTTG 386  
DB 181 GCGGTGATGATGCGAGCGCTGTCTGATGATGAGGACACCAAGCAGAGGCTCTTG 240  
QY 387 TTGTCTATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 446  
DB 241 TTGTCTATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
QY 447 TCATCTGACCGCCAGG 464  
DB 301 TCATCTGACCGCCAGG 318

RESULT 4  
US-09-815-343-1201  
Sequence 1201, Application US/09815343  
Patent No. US20010055596A1  
GENERAL INFORMATION:  
APPLICANT: Meagher, Madeline  
APPLICANT: Xu, Jiaqun  
APPLICANT: King, Gordon E.

TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER  
FILE REFERENCE: 210121.504  
CURRENT APPLICATION NUMBER: US/09/815,343  
CURRENT FILING DATE: 2001-03-22  
NUMBER OF SEQ ID NOS: 1556  
SOFTWARE: FASTSEQ for Windows Version 4.0  
SEQ ID NO 1201  
LENGTH: 318  
TYPE: DNA  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: misc-feature  
LOCATION: (1)...(318)  
OTHER INFORMATION: n = A,T,C or G  
US-09-815-343-1201

Query Match  
Best Local Similarity 38.2%; Score 317; DB 10; Length 318;  
Matches 317; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 147 CCTGAGAGCTGTGAGAGAGTTTGTCAATGCTCTACTACACCAACGATGATAGCGGCGC 206  
DB 1 CCTGAGAGCTGTGAGAGAGTTTGTCAATGCTCTACTACACCAACGATGATAGCGGCGC 60  
QY 207 GTTGTCTGTCCCGCTGTACATGGGACAGCCACCCCTGTCTGAAATGCAATGCTGTT 266  
DB 61 GTTGTCTGTCCCGCTGTACATGGGACAGCCACCCCTGTCTGAAATGCAATGCTGTT 120  
QY 267 CAGACAGAATCTCTGATGATGTTTGTAAATGTTCCCTTCCAGCGAGTTCCAAATCA 326  
DB 121 CAGACAGAATCTCTGATGATGTTTGTAAATGTTCCCTTCCAGCGAGTTCCAAATCA 180  
QY 327 GCGGTGATGATGCGAGCGCTGTCTGATGATGAGGACACCAAGCAGAGGCTCTTG 386  
DB 181 GCGGTGATGATGCGAGCGCTGTCTGATGATGAGGACACCAAGCAGAGGCTCTTG 240  
QY 387 TTGTCTATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 446  
DB 241 TTGTCTATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300  
QY 447 TCATCTGACCGCCAGG 464  
DB 301 TCATCTGACCGCCAGG 318

RESULT 5  
US-09-764-864-393  
Sequence 393, Application US/09764864  
Patent No. US20020132753A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT23  
CURRENT APPLICATION NUMBER: US/09/764,864  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1792  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 393  
LENGTH: 1387  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (1379)  
OTHER INFORMATION: n equals a,t,g, or c  
NAME/KEY: SITE  
LOCATION: (1380)  
OTHER INFORMATION: n equals a,t,g, or c  
US-09-764-864-393

Query Match 30.1%; Score 250; DB 10; Length 1387;

Best Local Similarity 71.0%; Pred. No. 1.9e-71;  
Matches 331; Conservative 0; Mismatches 135; Indels 0; Gaps 0;

OY 98 ATAGACAGAGATGGGATCTGTGTGATTTCAAGACCTATGTGATCAGGCGCTGCAGACT 157  
DB 52 ATAGGAGAGGCGGGGATGTGGGATTTTAAACTTATGTAGATCAGGACATGTAAGACT 111  
OY 158 GCTGAGAGATTTGATCATGTCTACTACACACATGGATAGCGGGCGCTTGTCTGCC 217  
DB 112 GCTGAGAGATTTGATCATGTCTACTACACATGGATAGCGGGCGCTTGTCTGCC 217  
OY 218 GCGCTGATCATGTGCGCACAGCCCTGCTGTGATGATGCGATGCTTTTCAGACAAGA 277  
DB 172 AGGCTGATCTGCGACAAGCCACTTATATGATGAAATGCTGTTTTCAGGCGCTGAT 231  
OY 278 TCCCTGATGATTTTGTGAAATGCTTCCCTTCCAGCATTTCAATACGCGTGTAGAC 337  
DB 232 GCGCTAAATATTTTGTGACACATTTGCTTGTAGTGTCCAGTCAATATGTAGAT 291  
OY 338 TGCCAGCCTGTCATGATGAGCCACCAAGCCAGACCGGCTTGTGTCTATCTGT 397  
DB 292 TGCCAGCCTGTCATGATGAGCCACCAAGCCAGACCGGCTTGTGTCTATCTGT 397  
OY 398 GGATCACTGAAGTTTGAAGGGGAAACAACAGGCGCTTCAACAGACCTTCACTGAC 457  
DB 352 GGAACCTGTGAAGTTTGAAGGGGAAACAACAGGCGCTTCAACAGACCTTCACTGAC 411  
OY 458 GCCCAGGCTTCAACCCACCAACAGTGTGAAAGTCGAAGTACCTTCCCTTCCAG 517  
DB 412 GCTGAGTCCACCTCCCAACATCTGTGTGAGATGCAAGTATGCTTCCCTTTCATCA 471  
OY 518 GACTGGCCAGCTAGTGGGGGGGAGAGCTCTCTTCTCTCATTC 563  
DB 472 GATGTGCTAGTGTATTAAGGGGCAAAAGTCCATCTCATTTGCTC 517

RESULT 6  
US-09-918-995-2901  
Sequence 2901, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918, 995  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235, 076  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 2901  
LENGTH: 415  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-918-995-2901

Query Match  
Best Local Similarity 99.1%; Pred. No. 7.7e-61;  
Matches 218; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 41 CCTGACGCGCCCTGTTCCCAAGCAGAGAAATACCTGTGAGCGCTTCTTCATA 100  
DB 196 CCGGCGACGCGCCCTGTTCCCAAGCAGAGAAATACCTGTGAGCGCTTCTTCATA 255  
OY 101 GAACAGAGATGATGCTGTGATTTCAAGACCTATGTGATCAGGCGCTGCAGACTGCT 160  
DB 256 GAACAGAGATGATGCTGTGATTTCAAGACCTATGTGATCAGGCGCTGCAGACTGCT 315  
OY 161 GAGGAGTTTGTCAATGTCTACTACACCAATGATAGCGGCGCTTGTCTGTCGCGC 220  
DB 316 GAGGAGTTTGTCAATGTCTACTACACCAATGATAGCGGCGCTTGTCTGTCGCGC 375

OY 221 CTGTACATGGCCACAGCCACCCCTGCTGTGATGATGCATG 260  
DB 376 CTGTACATGGCCACAGCCACCCCTGCTGTGATGATGCATG 415

RESULT 7  
US-09-918-995-6557  
Sequence 6557, Application US/09918995  
Publication No. US20030073623A1  
GENERAL INFORMATION:

APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918, 995  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US/09/235, 076  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 6557  
LENGTH: 417  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: (1)-(417)  
OTHER INFORMATION: n = A,T,C or G  
US-09-918-995-6557

Query Match  
Best Local Similarity 98.9%; Pred. No. 5.3e-46;  
Matches 173; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

OY 41 CCTGACGCGCCCTGTTCCCAAGCAGAGAAATACCTGTGAGCGCTTCTTCATA 100  
DB 242 CCGGCGACGCGCCCTGTTCCCAAGCAGAGAAATACCTGTGAGCGCTTCTTCATA 301  
OY 101 GAACAGAGATGATGCTGTGATTTCAAGACCTATGTGATCAGGCGCTGCAGACTGCT 160  
DB 302 GAACAGAGATGATGCTGTGATTTCAAGACCTATGTGATCAGGCGCTGCAGACTGCT 361  
OY 161 GAGGAGTTTGTCAATGTCTACTACACCAATGATAGCGGCGCTTGTCTGCT 215  
DB 362 GAGGAGTTTGTCAATGTCTACTACACCAATGATAGCGGCGCTTGTCTGCT 416

RESULT 8  
US-09-764-864-1682/c  
Sequence 1682, Application US/09764864  
Patent No. US20020132753A1  
GENERAL INFORMATION:

APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT223  
CURRENT APPLICATION NUMBER: US/09/764, 864  
PRIOR FILING DATE: 2001-01-17  
PRIOR APPLICATION data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1792  
SOFTWARE: Patenclin Ver. 2.0  
SEQ ID NO: 1682  
LENGTH: 8918  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-864-1682

Query Match  
Best Local Similarity 13.6%; Score 113; DB 10; Length 8918;  
Matches 152; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

OY 347 GTTCATGATGAAGCCACACCAAGCCAGCGTCTGTGATGATGATGATGATGATG 406  
DB 2199 GATGAAGAGCAAGCAAGCAAGCCCAACTGCTTCTTGTGACCAAGTGAAGTGTG 2140



QY 407 AAGTTGAGGGGGAACAACAGGAGGACTTCACCAAGACTTCATCCAGCCGCGCCAGGCC 466  
DB 2139 AAGTTGAGGGGGAACAACAGGAGGACTTCATCCAGGAGGACTTCGCTGAGTGGCTATCC 2080  
QY 467 TCACCCAGCAGCAACAGTGTGGAAGATCGAAGTGAAGTCTTCGCTTCAGAGAGTGGCC 526  
DB 2079 ACACCCAGCAACATCTGTGTGGAAGATTCGAAGGATGATGCTTCGCTTCAGAGATGGTCT 2020  
QY 527 AGCTAGTGGGGGTGCGAGAGGCTCTTTCCTTCATTC 563  
DB 2019 AGTAGTTAAGGGGCAAAAGCTCATCTCATTTGCTC 1983

RESULT 9  
US-09-764-1683/C  
Sequence 1683, Application US/09764864  
Patent No. US20020132753A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT23  
CURRENT APPLICATION NUMBER: US/09/764, 864  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1792  
SOFTWARE: Patentln Ver. 2.0  
SEQ ID NO 1683  
LENGTH: 8919  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-764-1683

Query Match 13.6%; Score 113; DB 10; Length 8919;  
Best Local Similarity 70.0%; Pred. No. 9.5e-26;  
Matches 152; Conservative 0; Mismatches 65; Indels 0; Gaps 0;

QY 347 GTTCATGATGAGACCCACACAGCCAGGAGGCTTCCTGTTGATCTGTGAGATCATG 406  
DB 2200 GATGAAGAGCAAGCAACTCATCTCCCAACTCATCTGTTGTGACAGAGTGAAGTGTG 2141  
QY 407 AAGTTGAGGGGGAACAACAGGAGGACTTCACCAAGACTTCATCCAGCCGCGCCAGGCC 466  
DB 2140 AAGTTGATGAGAAACAAACATCTTCACAGAGACTTCCTGCTGATCTCATGCTC 2081  
QY 467 TCACCCAGCAGCAACAGTGTGGAAGATCGAAGTGAAGTCTTCGCTTCAGAGAGTGGCC 526  
DB 2080 ACTCCAGCAACATCTGTGTGGAAGATTCGAAGTGAAGTCTTCGCTTCAGAGATGGTCT 2021  
QY 527 AGCTAGTGGGGGTGCGAGAGGCTCTTTCCTTCATTC 563  
DB 2020 AGTACTTAAGGGGCAAAAGCTCATCTCATTTGCTC 1984

RESULT 10  
US-09-918-995-36691/C  
Sequence 36691, Application US/09918995  
Publication No. US20030073633A1  
GENERAL INFORMATION:  
APPLICANT: Hyseq, Inc.  
TITLE OF INVENTION: NOVEL NUCLEIC ACID SEQUENCES OBTAINED  
FROM VARIOUS CDNA LIBRARIES  
FILE REFERENCE: 20411-756  
CURRENT APPLICATION NUMBER: US/09/918, 995  
CURRENT FILING DATE: 2001-07-30  
Prior application data removed - refer to PALM or file wrapper  
NUMBER OF SEQ ID NOS: 38054  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 36691  
LENGTH: 415  
TYPE: DNA  
ORGANISM: Homo sapiens

US-09-918-995-36691

Query Match 4.4%; Score 36.4; DB 9; Length 415;  
Best Local Similarity 52.7%; Pred. No. 0.23;  
Matches 79; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

QY 677 TTTTGTGCTCTAAGTCTAGAGAGTCCCTTCCTGAAATATATATCTGTTGATAGT 736  
DB 289 TTTCTGTGTTCTATTTTAT 230  
QY 737 TCCCTTCAAGAGTAACTTTCTATTTCTACTTGCAGAGAGTCTGATTC 796  
DB 229 TTTTGTGCTCTAAGTCTAGAGAGTCCCTTCCTGAAATATATATCTGTTGATAGT 170  
QY 797 GGAATCTGCAAT 826  
DB 169 CATATTCATCAAT 140

RESULT 11  
US-09-799-799-3  
Sequence 3, Application US/09799799  
Patent No. US20020132291A1  
GENERAL INFORMATION:  
APPLICANT: YE, Jane et al.  
TITLE OF INVENTION: ISOLATED HUMAN RAS-LIKE PROTEINS,  
NUCLEIC ACID MOLECULES ENCODING THESE HUMAN RAS-LIKE  
TITLE OF INVENTION: PROTEINS, AND USES THEREOF  
FILE REFERENCE: C1001157  
CURRENT APPLICATION NUMBER: US/09/799, 799  
CURRENT FILING DATE: 2001-03-07  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 3  
LENGTH: 89191  
TYPE: DNA  
ORGANISM: Human  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(89191)  
OTHER INFORMATION: n = A,T,C or G  
US-09-799-799-3

Query Match 4.4%; Score 36.4; DB 10; Length 89191;  
Best Local Similarity 50.6%; Pred. No. 7.6;  
Matches 88; Conservative 0; Mismatches 86; Indels 0; Gaps 0;

QY 621 GTTCATCTCTGTTGCGAGACATCGAGACTCCAGTGGCGAGTGAATCTT 680  
DB 7007 GTGACTGATTTTGTATCCCTGACCTGCTAATTCACATAGTCTGGAGTTTCT 7066  
QY 681 TGTGCTCAGTCTAGAGAGTCCCTTCCTAATATATATATATATATATATATATATAT 740  
DB 7067 TGTGCTCAGTCTAGAGAGTCCCTTCCTGATTTCTTGTGAGAGTCAATCTAT 7126  
QY 741 TTTCAAGTGAAT 794  
DB 7127 TTGCAACATGACAT 7180

RESULT 12  
US-09-764-877-3374  
Sequence 3374, Application US/09764877  
Patent No. US20020147140A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PC005  
CURRENT APPLICATION NUMBER: US/09/764, 877  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - refer to PALM or file wrapper  
NUMBER OF SEQ ID NOS: 4031  
SOFTWARE: Patentln Ver. 2.0

Query Match	4.2%	Score 34.8	DB 9	Length 1490
Best Local Similarity	58.8%	Pred. No. 1.8		
Matches	60	Conservative	0	Mismatches 42
			Indels	0
			Gaps	0
471	CGACGACCACTGTGGGAAGATCGCAAGTCACTGCTTCGCTTCGAGGACTGGCCACT	530		
236	CCAGTCTCTGGGCACAGCGCGGTGGAGCGCTGCTCTTGTGTGACGACGACAGAGA	177		

Mon Jun 30 08:51:11 2003

us-09-763-902b-10.rnpb

Page 7

Db 176 TGAGGGGAGGAGGATTCTACGATGCAAGGCACCCGTAC 135

Search completed: June 28, 2003, 03:41:58  
Job time : 142.193 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2003, 00:27:44 ; Search time 45.3644 Seconds  
(Without alignments)  
5611.049 Million cell updates/sec

Title: US-09-763-902B-10

Perfect score: 830

Sequence: 1 tgaacttgcatcactg.....ataattaataacacatg 830

Scoring table: IDENTITY\_NUC

Gapop 10.0, Gapext 1.0

Searched: 441362 seqs, 15338381 residues

Total number of hits satisfying chosen parameters: 882724

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database: Issued Patents, NA:

1: /cgn2\_6/ptodata/1/lna/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/lna/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/lna/6A.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/lna/6B.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/lna/PCITUS.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/lna/Backfiles1.seq:\*

Pred. NO. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	41.2	5.0	7218 1	US-08-232-463-14 Sequence 14, Appl
2	35	4.2	8478 3	US-08-676-818-1 Sequence 1, Appl
3	35	4.2	8478 4	US-09-407-549-1 Sequence 1, Appl
4	33.6	4.0	3957 1	US-07-689-008-5 Sequence 5, Appl
5	33.6	4.0	9540 1	US-07-689-008-5 Sequence 1, Appl
6	33.6	4.0	4403765 4	US-09-103-840A-2 Sequence 2, Appl
7	33.6	4.0	4411529 4	US-09-103-840A-1 Sequence 1, Appl
8	33	4.0	2861 4	US-09-134-001C-1659 Sequence 1659, Ap
9	32	3.9	11811 4	US-09-078-294-7 Sequence 7, Appl
10	31.4	3.8	16836 4	US-09-147-236-1 Sequence 10, Appl
11	31.4	3.8	16836 4	US-09-147-236-1 Sequence 10, Appl
12	31.2	3.8	3114 3	US-08-946-026-12 Sequence 12, Appl
13	30.8	3.7	2109 2	US-08-835-099A-5 Sequence 5, Appl
14	30.8	3.7	2109 2	US-08-835-099A-5 Sequence 5, Appl
15	30.8	3.7	2136 2	US-08-835-099A-6 Sequence 6, Appl
16	30.8	3.7	2136 3	US-09-157-349-6 Sequence 6, Appl
17	30.8	3.7	2237 1	US-08-487-135B-1 Sequence 1, Appl
18	30.8	3.7	2237 2	US-08-915-972A-1 Sequence 1, Appl
19	30.8	3.7	2237 2	US-09-177-909-1 Sequence 1, Appl
20	30.6	3.7	1179 4	US-09-247-373B-43 Sequence 43, Appl
21	30.6	3.7	2277 1	US-08-676-967-2 Sequence 2, Appl
22	30.6	3.7	2277 1	US-08-676-967-2 Sequence 2, Appl
23	30.6	3.7	2277 2	US-09-098-487-2 Sequence 2, Appl
24	30.4	3.7	848 4	US-08-905-223-27 Sequence 27, Appl
25	30.4	3.7	848 4	US-09-247-155-27 Sequence 27, Appl
26	30.4	3.7	1364 1	US-08-265-087-3 Sequence 3, Appl
27	30.4	3.7	1364 1	US-08-621-493-3 Sequence 3, Appl

C 28	30.4	3.7	1364 2	US-08-965-688-3 Sequence 3, Appl
C 29	30.4	3.7	1364 4	US-09-260-173-3 Sequence 3, Appl
C 30	30.4	3.7	4601 1	US-08-470-058-3 Sequence 3, Appl
C 31	30.4	3.7	4601 3	US-09-037-188-3 Sequence 3, Appl
C 32	30.4	3.7	4601 4	US-09-285-310-3 Sequence 3, Appl
C 33	30.4	3.7	5095 1	US-08-092-817-3 Sequence 3, Appl
C 34	30.4	3.7	5095 4	US-08-485-128-3 Sequence 3, Appl
C 35	30.2	3.6	55827 4	US-09-813-133A-3 Sequence 454, App
C 36	30	3.6	565 4	US-09-328-111-454 Sequence 63, Appl
C 37	30	3.6	1114 4	US-09-378-088A-63 Sequence 10, Appl
C 38	30	3.6	1158 3	US-08-633-993A-10 Sequence 10, Appl
C 39	30	3.6	1158 4	US-08-844-188-10 Sequence 10, Appl
C 40	30	3.6	1158 4	US-09-378-088A-10 Sequence 67, Appl
C 41	30	3.6	1158 4	US-09-378-088A-67 Sequence 89, Appl
C 42	30	3.6	1186 4	US-09-378-088A-89 Sequence 11, Appl
C 43	30	3.6	1334 4	US-09-347-798-11 Sequence 30, Appl
C 44	30	3.6	2015 3	US-08-633-993A-30 Sequence 30, Appl
C 45	30	3.6	2015 3	US-08-844-188-30 Sequence 30, Appl

## ALIGNMENTS

RESULT 1  
US-08-232-463-14  
Sequence 14, Application US/08232463  
Patent No. 5670367  
GENERAL INFORMATION:  
APPLICANT: DORNER, F.  
APPLICANT: SCHEIFLINGER, F.  
APPLICANT: FALKNER, F. G.  
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
NUMBER OF SEQUENCES: 52  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 1800 Diagonal Road, Suite 500  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-0299  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/232,463  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US/07/935,313  
FILING DATE:  
APPLICATION NUMBER: EP 91 114 300.6  
FILING DATE: 26-AUG-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: BENT, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703)836-9300  
TELEFAX: (703)683-4109  
TELEX: 899149  
INFORMATION FOR SEQ ID NO: 14:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 7218 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
IMMEDIATE SOURCE:  
CLONE: pTZgpt-Fls  
US-08-232-463-14  
Query Match 5.0%; Score 41.2; DB 1; Length 7218;

[illegible]

OY 208 TTTCCTCCCGCTGTACATGGGACAGCCCGTGTGATGGCAATGCTTC 267  
DB 3422 TTATTTTCTATATGACATGACAGACAGTGTGAATATGGTTATGTGGTAA 3481  
OY 268 AGACAGAAATCCTTGAGT 286  
DB 3482 AGGAGATCAGACACAGCAAT 3500

## RESULT 4

US-07-689-008-5/c  
Sequence 5, Application US/07689008  
Patent No. 5268274

## GENERAL INFORMATION:

APPLICANT: Ben-Bassat, Arie  
APPLICANT: Calhoon, Roger D  
APPLICANT: Fear, Anna L  
APPLICANT: Gelfand, David H  
APPLICANT: Meade, James H  
APPLICANT: Tal, Rony  
APPLICANT: Wong, Hing  
APPLICANT: Benziman, Moshe  
TITLE OF INVENTION: METHODS AND NUCLEIC ACID SEQUENCES FOR THE  
NUMBER OF SEQUENCES: 15  
TITLE OF INVENTION: EXPRESSION OF CELLULOSE SYNTHASE OPERON  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McCutchen, Doyle, Brown & Enersen  
STREET: Three Embarcadero Center  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/689,008  
FILING DATE: 19910422  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 337,194  
FILING DATE: 12-APR-1989  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 496,236  
FILING DATE: 23-MAR-1990

## ATTORNEY/AGENT INFORMATION:

NAME: Murphy, Lisabeth Felix  
REGISTRATION NUMBER: 31547  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 393-2000  
TELEFAX: (415) 393-2286  
TELEX: 340817 MACPAG SFO

## INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:  
LENGTH: 3957 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-07-689-008-5

## Query Match

Best Local Similarity 4.0%; Score 33.6; DB 1; Length 3957;  
Matches 63; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

OY 468 CACCCAGCAACAGTGTGGAAGATCGCAAGTACTGCTTCGCTTCAGAGACTGGGCA 527  
DB 484 CACCGTGGAAACAGCTTCTGTGTAGCCCGCACCGCGTCTGCGCCGCAACCGCA 425  
OY 528 GCTAGTGGGGGTGGCAGAGGCTCTTTGCTTCAATGACCCCTAGCTCTGTAG 579

DB 424 GCGAGCGATGTGCGACAGTCCCTTGGAGATGCGCCGCTCCGTAGACAG 373

## RESULT 5

US-07-689-008-1/c  
Sequence 1, Application US/07689008  
Patent No. 5268274

## GENERAL INFORMATION:

APPLICANT: Ben-Bassat, Arie  
APPLICANT: Calhoon, Roger D  
APPLICANT: Fear, Anna L  
APPLICANT: Gelfand, David H  
APPLICANT: Meade, James H  
APPLICANT: Tal, Rony  
APPLICANT: Wong, Hing  
APPLICANT: Benziman, Moshe  
TITLE OF INVENTION: METHODS AND NUCLEIC ACID SEQUENCES FOR THE  
NUMBER OF SEQUENCES: 15  
TITLE OF INVENTION: EXPRESSION OF CELLULOSE SYNTHASE OPERON  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: McCutchen, Doyle, Brown & Enersen  
STREET: Three Embarcadero Center  
CITY: San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94111

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.24  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/689,008  
FILING DATE: 19910422  
CLASSIFICATION: 435  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 337,194  
FILING DATE: 12-APR-1989  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 496,236  
FILING DATE: 23-MAR-1990

## ATTORNEY/AGENT INFORMATION:

NAME: Murphy, Lisabeth Felix  
REGISTRATION NUMBER: 31547  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 393-2000  
TELEFAX: (415) 393-2286  
TELEX: 340817 MACPAG SFO

## INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:  
LENGTH: 9540 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: join(328..2589, 2594..4999, 5005..8961, 8964  
LOCATION: 9431)  
OTHER INFORMATION:  
US-07-689-008-1

## Query Match

Best Local Similarity 4.0%; Score 33.6; DB 1; Length 9540;  
Matches 63; Conservative 0; Mismatches 49; Indels 0; Gaps 0;

OY 468 CACCCAGCAACAGTGTGGAAGATCGCAAGTACTGCTTCGCTTCAGAGACTGGGCA 527  
DB 5488 CACCGTGGAAACAGCTTCTGTGTAGCCCGCACCGCGTCTGCGCCGCAACCGCA 5429  
OY 528 GCTAGTGGGGGTGGCAGAGGCTCTTTGCTTCAATGACCCCTAGCTCTGTAG 579

Db 5428 GCGAGCGATCTGCGACAGGTGCTTGGAGATGCGCCGCTGCTGAGCAG 5377

## RESULT 6

US-09-103-840A-2

Sequence 2, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 2

LENGTH: 4403765

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

FEATURE:

OTHER INFORMATION: CDC 1551

OTHER INFORMATION: "n" bases at various positions throughout the sequence

US-09-103-840A-2

Query Match

Best Local Similarity 4.0%; Score 33.6; DB 4; Length 4403765;

Matches 81; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Db 400 ATGAGGAACTTGTAGGGGACAAACAGGAGCTTACACCACTTCTGACCGC 459

Db 2768205 ATCGGTGGGTGATGGCTGTATCCGCACCTGGACATCAGCAGGACGACCGCGC 2768264

Db 460 CCAGGCTTACCCAGCAGACAGTGTGAAGATCGCAGAGTGTCTCCGTTCCAGCA 519

Db 2768265 CGAGGAGTCTCTTACGAGACCCGAGGTTCCGCCGATTTGCGGCTGTCTGAGGG 2768324

Db 520 CTGGGCCAGCTAGTGGGGTGCAGAGTCTCTTCTTC 559

Db 2768325 CCAGGCCGCGGTGACGATCGTTGGGGGCCCACTTC 2768364

RESULT 7

US-09-103-840A-1

Sequence 1, Application US/09103840A

Patent No. 6294328

GENERAL INFORMATION:

APPLICANT: FLEISCHMAN, Robert D.

APPLICANT: WHITE, Owen R.

APPLICANT: FRASER, Claire M.

APPLICANT: VENTER, John C.

TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM

FILE REFERENCE: 24366-20007.00

CURRENT APPLICATION NUMBER: US/09/103,840A

CURRENT FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 2

SOFTWARE: Patentin Ver. 2.1

SEQ ID NO 1

LENGTH: 4411529

TYPE: DNA

ORGANISM: Mycobacterium tuberculosis

OTHER INFORMATION: H37RV

US-09-103-840A-1

Query Match

Best Local Similarity 4.0%; Score 33.6; DB 4; Length 4411529;

Matches 81; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

Db 400 ATGAGTGAAGTTGTAGGGGACAAACAGGAGCTTACACCACTTCTGACCGC 459

Db 2771395 ATCGGTGGGTGATGGCTGTATCCGCACCTGGACATCAGCAGGACGACCGCGC 2771454

Db 460 CCAGGCTTACCCAGCAGACAGTGTGAAGATCGCAGAGTGTCTCCGTTCCAGCA 519

Db 2771455 CGAGGAGTCTCTTACGAGACCCGAGGTTCCGCCGATTTGCGGCTGTCTGAGGG 2771514

Db 520 CTGGGCCAGCTAGTGGGGTGCAGAGTCTCTTCTTC 559

Db 2771515 CCAGGCCGCGGTGACGATCGTTGGGGGCCCACTTC 2771554

RESULT 8

US-09-134-001C-1659/c

Sequence 1659, Application US/09134001C

Patent No. 6380370

GENERAL INFORMATION:

APPLICANT: Lynn Doucette-Stamm et al

TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC

FILE REFERENCE: GTC-007

CURRENT APPLICATION NUMBER: US/09/134,001C

CURRENT FILING DATE: 1997-11-08

PRIOR APPLICATION NUMBER: US 60/064,964

PRIOR FILING DATE: 1997-08-14

NUMBER OF SEQ ID NOS: 5674

SEQ ID NO 1659

LENGTH: 2661

TYPE: DNA

ORGANISM: Staphylococcus epidermidis

US-09-134-001C-1659

Query Match

Best Local Similarity 4.0%; Score 33; DB 4; Length 2661;

Matches 57; Conservative 0; Mismatches 40; Indels 0; Gaps 0;

Db 672 AACTCTTTTGTGCTCAAGTCTAGAGTCCCTTCCGATATATATCTGTTGCA 731

Db 2253 AACTATCATATATATATATCTGTTGTTCTTTTCAATATGTTATTTAATC 2194

Db 732 TACTTCTTTTCAAGTATTAACCTTTCTATTTT 768

Db 2193 TATCGTCTTTTAAGTTGTAATTTCTTCTACTTT 2157

RESULT 9

US-09-078-294-7

Sequence 7, Application US/09078294

Patent No. 6265211

GENERAL INFORMATION:

APPLICANT: Cho, Kong-Hong Andy

APPLICANT: Du Sait, Desiree

APPLICANT: Cancilla, Michael R.

TITLE OF INVENTION: A NOVEL NUCLEIC ACID MOLECULE

FILE REFERENCE: Davles Col

CURRENT APPLICATION NUMBER: US/09/078,294

CURRENT FILING DATE: 1998-05-13

NUMBER OF SEQ ID NOS: 29

SOFTWARE: Patentin Ver. 2.0

SEQ ID NO 7

LENGTH: 11811

TYPE: DNA

ORGANISM: BAC-12 contig 3

US-09-078-294-7

Query Match

Best Local Similarity 3.9%; Score 32; DB 4; Length 11811;

Matches 59; Conservative 0; Mismatches 45; Indels 0; Gaps 0;

Db 671 GAATCTTTTGTGCTCAAGTTCTAGAGTCCCTTCCGATATATATCTGTTGTC 730



DB 4379 GCACCTGGTACTGATCTTTAATTTTGTACTTAACTAGTGGGCAATATCTGTATTT 4438  
QY 731 ATGTTTCCTTTCAAGTAGTAACCTTTCTATTTTCTACTT 774  
DB 4439 ATGTTTACTTTTATTTTCTTATGATTAATAGTTGACGAT 4482

RESULT 10  
US-09-147-236-1/c  
Sequence 1, Application US/09147236A  
Patent No. 6316251  
GENERAL INFORMATION:

APPLICANT: TONOUCHI, Naoto  
APPLICANT: TSUCHIDA, Takayasu  
APPLICANT: YOSHINAGA, Fumihito  
APPLICANT: TAHARA, Naoki  
APPLICANT: HAYASHI, Takahisa  
TITLE OF INVENTION: NOVEL GENE, GROUP OF GENES, AND NOVEL BETA-GLUCOSIDASE  
FILE REFERENCE: 6537-011-0PCT  
CURRENT FILING DATE: 1999-04-08  
CURRENT APPLICATION NUMBER: PCT/JP97/03633  
EARLIER FILING DATE: 1997-10-09  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 1  
LENGTH: 16836  
TYPE: DNA  
ORGANISM: Acetobacter xylinum

FEATURE:  
NAME/KEY: CDS  
LOCATION: (869)..(1891)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (3101)..(5368)  
FEATURE:

NAME/KEY: CDS  
LOCATION: (5373)..(7778)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (7784)..(11761)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (11764)..(12231)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (12448)..(14652)  
FEATURE:

OTHER INFORMATION: n at positions 15741 and 15767 may be a, g, t, or  
OTHER INFORMATION: c  
US-09-147-236-1

Query Match 3.8%; Score 31.4; DB 4; Length 16836;  
Best Local Similarity 54.9%; Pred. No. 22;  
Matches 62; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 468 CACCAGACACAGTGTGGAAGATGCAAGTACTCTTCGCTTCACAGACTGGGCA 527  
DB 8267 CACCGTGAACAGCTTCTGTCACCCGCCGCTGCTGCGCCGGAACCGCCA 8208  
QY 528 GCTAGTGGGGGTGGCAGAGGCTCTTTGCTTATTCAGCCCTAGCTGTAGA 580  
DB 8207 GCGAACGATCTGTGACAGAGTGTGCTTTGGGAATGGCCGCTGAGACAGA 8155

RESULT 11

US-09-147-236-10/c  
Sequence 10, Application US/09147236A  
Patent No. 6316251  
GENERAL INFORMATION:  
APPLICANT: TONOUCHI, Naoto  
APPLICANT: TSUCHIDA, Takayasu

APPLICANT: YOSHINAGA, Fumihito  
APPLICANT: TAHARA, Naoki  
APPLICANT: HAYASHI, Takahisa  
TITLE OF INVENTION: NOVEL GENE, GROUP OF GENES, AND NOVEL BETA-GLUCOSIDASE  
FILE REFERENCE: 6537-011-0PCT  
CURRENT FILING DATE: 1999-04-08  
CURRENT APPLICATION NUMBER: PCT/JP97/03633  
EARLIER FILING DATE: 1997-10-09  
NUMBER OF SEQ ID NOS: 12  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 10  
LENGTH: 16836  
TYPE: DNA  
ORGANISM: Acetobacter xylinum

FEATURE:  
NAME/KEY: CDS  
LOCATION: (1891)..(2922)  
FEATURE:

OTHER INFORMATION: Nucleotide sequence is the same as SEQ ID NO:1  
OTHER INFORMATION: t  
US-09-147-236-10

Query Match 3.8%; Score 31.4; DB 4; Length 16836;  
Best Local Similarity 54.9%; Pred. No. 22;  
Matches 62; Conservative 0; Mismatches 51; Indels 0; Gaps 0;

QY 468 CACCAGACACAGTGTGGAAGATGCAAGTACTCTTCGCTTCACAGACTGGGCA 527  
DB 8267 CACCGTGAACAGCTTCTGTCACCCGCCGCTGCTGCGCCGGAACCGCCA 8208  
QY 528 GCTAGTGGGGGTGGCAGAGGCTCTTTGCTTATTCAGCCCTAGCTGTAGA 580  
DB 8207 GCGAACGATCTGTGACAGAGTGTGCTTTGGGAATGGCCGCTGAGACAGA 8155

RESULT 12

US-08-946-026-12  
Sequence 12, Application US/08946026  
Patent No. 6034218  
GENERAL INFORMATION:

APPLICANT: Reed, Steven G.  
APPLICANT: Dillion, Davin C.  
APPLICANT: Twartzik, Daniel R.  
APPLICANT: Mitcham, Jennifer L.  
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR IMMUNOTHERAPY  
TITLE OF INVENTION: AND IMMUNODIAGNOSIS OF PROSTATE CANCER  
NUMBER OF SEQUENCES: 59  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: SEED AND BERRY LLP  
STREET: 6300 Columbia Center, 701 Fifth Avenue  
CITY: Seattle  
STATE: Washington  
COUNTRY: USA  
ZIP: 98104-7092

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/946, 026  
FILING DATE: 07-OCT-1997  
CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: MAKI, David J.  
REGISTRATION NUMBER: 31,392  
REFERENCE/DOCKET NUMBER: 210121.424C1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (206) 622-4900  
TELEFAX: (206) 682-6031

## INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:  
LENGTH: 3114 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-946-026-12

Query Match  
Best Local Similarity 58.7%; Score 31.2; DB 3; Length 3114;  
Matches 54; Conservative 0; Mismatches 38; Indels 0; Gaps 0;

QY 204 GGGGTTGCTGTCGCCGCTGTACATGGGACACACCCGCTGTGATGCAATGCTG 263  
DB 2581 GTGTGTTCTGCTGCTCCCTGTGTCAGGCTATCTGTAGGGTGACCTGGAATTCGAATTCG 2640  
QY 264 TTTCAGGACAGAACTTCCTGAGTGAAGTTT 295  
DB 2641 TTTCCTTGTAAATATTTGTCTGCTCTTT 2672

## RESULT 13

US-08-835-099A-5  
Sequence 5, Application US/08835099A  
Patent No. 5874277

## GENERAL INFORMATION:

APPLICANT: SHINTANI, Yasushi  
APPLICANT: NISHI, Kazuori  
APPLICANT: KAWAMOTO, Tomohiro  
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION  
TITLE OF INVENTION: AND USE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP  
STREET: 130 Water Street  
CITY: Boston  
STATE: MA

COUNTRY: USA  
ZIP: 02109

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/835,099A  
FILING DATE: 04-APR-1997

CLASSIFICATION: 514  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 083649/1996  
FILING DATE: 05-APR-1996

APPLICATION NUMBER: 97105508.2  
FILING DATE: 03-APR-1997

ATTORNEY/AGENT INFORMATION:  
NAME: Resnick, David S

REGISTRATION NUMBER: 34,235  
REFERENCE/DOCKET NUMBER: 47342

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440  
TELEX:

INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:

LENGTH: 2109 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

MOLECULE TYPE: CDNA  
US-08-835-099A-5

Query Match  
Best Local Similarity 57.7%; Score 30.8; DB 2; Length 2109;  
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 374 ACCAGGTCCTGTTGTATCTGTGATGATGAGTTTGAGGAGCAACACAGGAC 433  
DB 1288 AGCATGGGCTATGCGGCTACACAGGTTCACAGAGAGTGAGATCAGGAGCACAC 1347

QY 434 TTCAACGACAGATTCTATCTGACCGCCAGGCTCAGCCAGCAACA 479  
DB 1348 TTGGGCGGGATTTCTCTGCGCTACACAGCCCTCAGCCGCAACA 1393

## RESULT 14

US-09-157-349-5  
Sequence 5, Application US/09157349  
Patent No. 6068990

## GENERAL INFORMATION:

APPLICANT: SHINTANI, Yasushi  
APPLICANT: NISHI, Kazuori  
APPLICANT: KAWAMOTO, Tomohiro  
TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION  
TITLE OF INVENTION: AND USE  
NUMBER OF SEQUENCES: 18  
CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP  
STREET: 130 Water Street  
CITY: Boston  
STATE: MA

COUNTRY: USA  
ZIP: 02109

## COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS

SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/157,349  
FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/835,099  
FILING DATE:

APPLICATION NUMBER: 97105508.2  
FILING DATE: 03-APR-1997

ATTORNEY/AGENT INFORMATION:  
NAME: Resnick, David S

REGISTRATION NUMBER: 34,235  
REFERENCE/DOCKET NUMBER: 47342

TELECOMMUNICATION INFORMATION:  
TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440  
TELEX:

INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:

LENGTH: 2109 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear

MOLECULE TYPE: CDNA  
US-09-157-349-5

Query Match  
Best Local Similarity 55.7%; Score 30.8; DB 3; Length 2109;  
Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

QY 374 ACCAGGTCCTGTTGTATCTGTGATGATGAGTTTGAGGAGCAACACAGGAC 433  
DB 1288 AGCATGGGCTATGCGGCTACACAGGTTCACAGAGAGTGAGATCAGGAGCACAC 1347

QY 434 TTCAACGACAGATTCTATCTGACCGCCAGGCTCAGCCAGCAACA 479  
DB 1348 TTGGGCGGGATTTCTCTGCGCTACACAGCCCTCAGCCGCAACA 1393

## RESULT 15

US-08-835-099A-6

; Sequence 6, Application US/08835099A  
Patent No. 5874277

GENERAL INFORMATION:

APPLICANT: SHINTANI, Yasushi

APPLICANT: NISHI, Kazuo

APPLICANT: KANAMOTO, Tomohiro

TITLE OF INVENTION: NOVEL PROTEINS, THEIR PRODUCTION

TITLE OF INVENTION: AND USE

NUMBER OF SEQUENCES: 18

CORRESPONDENCE ADDRESS:

ADDRESSEE: DIKE, BRONSTEIN, ROBERTS & CUSMAN, LLP

STREET: 130 Water Street

City: Boston

STATE: MA

COUNTRY: USA

ZIP: 02109

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette

COMPUTER: IBM compatible

OPERATING SYSTEM: DOS for Windows Version 2.0

SOFTWARE: FASTSEQ for Windows Version 2.0

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/835,099A

FILING DATE: 04-APR-1997

CLASSIFICATION: 514

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 083649/1996

FILING DATE: 05-APR-1996

APPLICATION NUMBER: 97105508.2

FILING DATE: 03-APR-1997

ATTORNEY/AGENT INFORMATION:

NAME: Resnick, David S

REGISTRATION NUMBER: 34,235

REFERENCE/DOCKET NUMBER: 47342

TELECOMMUNICATION INFORMATION:

TELEPHONE: 617-523-3400

TELEFAX: 617-523-6440

TELEX:

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 2136 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: CDNA

US-08-835-099A-6

Query Match 3.7%; Score 30.8; DB 2; Length 2136;

Best Local Similarity 55.7%; Pred. NO. 10;

Matches 59; Conservative 0; Mismatches 47; Indels 0; Gaps 0;

OY	374	ACCACGCTCCTGTTGTCATCTGTGATCACTGAGTTTGAGGGGACAAACAGCGGAC	433
DB	1315	AGCATCGGCTATGCGCTGTACCAAGAGCTGCCAAGAGAGAGATCACACGAGACGAC	1374
OY	434	TTCAACCAAGAACTCATCTGACCGCCGAGGCTCACCACACACA	479
DB	1375	TTGGCCGGGATTTCTCTGCGCTACCAAGCCCTCAGCCCGCACCA	1420

Search completed: June 28, 2003, 03:36:22  
Job time : 53.3644 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2003, 00:22:51 ; Search time 224.7 Seconds

(without alignments)  
10761.067 Million cell updates/sec

Title: US-09-763-902b-10

Perfect score: 830  
Sequence: 1 tgaacttgcatcattcactg.....ataattataataacacatg 830

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 segs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: 1: gb\_ba: 2: gb\_hg: 3: gb\_in: 4: gb\_om: 5: gb\_ov: 6: gb\_pat: 7: gb\_ph: 8: gb\_pl: 9: gb\_pr: 10: gb\_ro: 11: gb\_sts: 12: gb\_sy: 13: gb\_un: 14: gb\_vl: 15: em\_ba: 16: em\_fun: 17: em\_hum: 18: em\_in: 19: em\_mu: 20: em\_om: 21: em\_or: 22: em\_ov: 23: em\_pat: 24: em\_ph: 25: em\_pl: 26: em\_ro: 27: em\_sts: 28: em\_un: 29: em\_vl: 30: em\_hg\_hum: 31: em\_hg\_inv: 32: em\_hg\_other: 33: em\_hg\_mus: 34: em\_hg\_pin: 35: em\_hg\_rod: 36: em\_hg\_mam: 37: em\_hg\_vrt: 38: em\_sy: 39: em\_hgo\_hum: 40: em\_hgo\_mus: 41: em\_hgo\_other:

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	830	100.0	112180	9	HSJ322G13	AL096677 Human DNA
2	786.8	94.8	1176	9	BC003410	BC003410 Homo sapi
3	759.8	91.5	965	9	AF156957	AF156957 Homo sapi
4	736.4	88.7	939	9	BC002687	BC002687 Homo sapi
5	733.2	88.3	963	9	BC003029	BC003029 Homo sapi
6	733.2	88.3	972	9	BC000759	BC000759 Homo sapi
7	719.2	86.7	1049	9	AK026360	AK026360 Homo sapi
8	508.4	61.3	135619	9	AL354809	AL354809 Human DNA
9	508.4	61.3	147310	2	AL513495	AL513495 Homo sapi
10	492	59.3	186415	2	AC110699	AC110699 Rattus no
11	381.2	45.9	644	10	AF156958	AF156958 Mus muscu
12	357.8	43.1	182118	2	AC128310	AC128310 Rattus no
13	317	38.2	318	6	AX261550	AX261550 Sequence
14	287.6	34.7	153483	2	AC111510	AC111510 Rattus no
15	251.6	30.3	896	9	AF201942	AF201942 Homo sapi
16	251.6	30.3	2577	9	BC014888	BC014888 Homo sapi
17	251.6	30.3	2692	9	AK023289	AK023289 Homo sapi
18	250.2	30.1	429	9	AF246127	AF246127 Homo sapi
19	250.2	30.1	429	9	HS277591	HS277591 Homo sapi
20	248.6	26.9	567	9	HS277591	HS277591 Homo sapi
21	223	24.7	6	AX330381	AX330381 Sequence	
22	132	15.9	133	11	G42938	G42938 WtAF-410-ST
23	113	13.6	17863	9	HS596C15	AL031387 Human DNA
24	105	12.7	148320	2	AC107438	AC107438 Rattus no
25	105	12.7	173939	2	AC110357	AC110357 Rattus no
26	105	12.7	186222	2	AC123243	AC123243 Rattus no
27	105	12.7	232171	10	AL731672	AL731672 Mouse DNA
28	81.2	9.8	238817	2	AC102111	AC102111 Mus muscu
29	76.4	9.2	507	3	AF156959	AF156959 Drosophill
30	76.4	9.2	538	3	AY071250	AY071250 Drosophill
31	59.4	7.2	295166	2	AL161914	AL161914 Homo sapi
32	58.8	7.1	42244	2	AC020482	AC020482 Drosophill
33	58.8	7.1	147086	3	AC009182	AC009182 Drosophill
34	58.8	7.1	175118	3	AC010842	AC010842 Drosophill
35	58.8	7.1	303823	3	AE003462	AE003462 Drosophill
36	57.6	6.9	173939	2	AC110357	AC110357 Rattus no
37	56	6.7	203822	2	AC073698	AC073698 Mus muscu
38	48.6	5.9	207621	2	AL845495	AL845495 Mus muscu
39	47.4	5.7	175991	2	AC094606	AC094606 Rattus no
40	41.6	5.0	38979	2	AC104197	AC104197 Mus muscu
41	41.2	5.0	7218	6	166494	166494 Sequence 14
42	41.2	5.0	110000	9	AL355364	AL355364 Homo sapi
43	41.2	5.0	125020	9	AF429315	AF429315 Homo sapi
44	41.2	5.0	139325	9	AL589943	AL589943 Human DNA
45	41	4.9	215219	2	AC118225	AC118225 Mus muscu

# ALIGNMENTS

RESULT 1  
HSJ322G13  
LOCUS  
DEFINITION  
Human DNA sequence from clone RP3-322G13 on chromosome 20p11.21-12.3 Contains the gene for NTF2-related export protein (NTF1), a gene for zinc finger protein FLJ21794, two putative novel genes, a gene for beta-soluble NSF attachment protein (SNAP-beta), a novel gene for a protein similar to cystatin, another novel gene for a protein similar to cystatin 8 (CST8), ESTs, STSS, GSSs and six Cpg islands, complete sequence.  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE

ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Metazoa; Primates; Catarrhini; Hominiidae; Homo.  
 REFERENCE 1 (bases 1 to 112180)  
 AUTHORS Lovell, J.  
 TITLE Direct Submission  
 JOURNAL Submitted (13-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 On Oct 17, 2000 this sequence version replaced gi:9795218.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 The following abbreviations are used to associate primary accession  
 numbers given in the feature table with their source databases:  
 Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information  
 on the WORMPEP database can be found at  
 http://www.sanger.ac.uk/projects/C\_elegans/wormpep This sequence  
 was generated from part of bacterial clone contigs of human  
 chromosome 20, constructed by the Sanger Centre Chromosome 20  
 Mapping Group. Further information can be found at  
 http://www.sanger.ac.uk/HGP/Chr20  
 This sequence is the entire insert of clone RP3-322G13. The true  
 left end of clone RP3-333B15 is at 28739 in this sequence. The true  
 right end of clone RP5-851M4 is at 70589 in this sequence. This  
 sequence was finished as follows unless otherwise noted: all  
 regions were either double-stranded or sequenced with an alternate  
 chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
 as compressions and repeats; all regions were covered by at least  
 one plasmid subclone or more than one M13 subclone; and the  
 assembly was confirmed by restriction digest. RP3-322G13 is from  
 the library RPCI-3 constructed by the group of Pieter de Jong. For  
 further details see  
 http://www.choil.org/bacpac/home.htm  
 VECTOR: pCYPAC2.

FEATURES  
 Source 1. 112180  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="20"  
 /map="p11.21-12.3"  
 /clone="RP3-322G13"  
 /clone\_id="RPCI-3"  
 79. 144  
 /note="LTR16C repeat: matches 321. .387 of consensus"  
 790. 1085  
 /note="AluSg repeat: matches 1. .300 of consensus"  
 1247. 1357  
 /note="LTR1H repeat: matches 384. .532 of consensus"  
 1748. 1831  
 /note="21 copies 4 mer ggtg 69% conserved"  
 1757. 1826  
 /note="5 copies 14 mer 75% conserved"  
 complement(join(2155. 2811,6544. .6700))  
 /gene="dJ322G13.7"  
 /gene="dJ322G13.7"  
 complement(join(<2155. .2811,6544. .6700))  
 /product="dJ322G13.7 (putative novel transcript)"  
 /note="match: ESTs: Em:BE891551"  
 /evidence=not\_experimental  
 2488. 2527  
 /note="MIR repeat: matches 108. .147 of consensus"  
 3133. 3309  
 /note="MIR repeat: matches 76. .254 of consensus"  
 3727. 4165  
 /note="LTR2 repeat: matches 1. .449 of consensus"  
 4279. 4345  
 repeat\_region  
 /note="12 repeat: matches 2679. .2745 of consensus"  
 5214. 6623  
 misc\_feature

gene  
 join(6067. 6275,9195. .9991)  
 /gene="dJ322G13.1"  
 join(6067. 6275,9195. .9991)  
 /gene="dJ322G13.1"  
 /product="dJ322G13.1 (NMF2-related export protein NXT1)"  
 /note="match: CDNAS: Em:AF156957  
 match: ESTs: Em:R35341 Em:AA223265 Em:AM249178"  
 /evidence=not\_experimental  
 7047. 7168  
 /note="MIR repeat: matches 99. .220 of consensus"  
 7169. 7507  
 /note="MER2 repeat: matches 1. .341 of consensus"  
 7508. 7523  
 /note="MIR repeat: matches 220. .235 of consensus"  
 8454. 8548  
 /note="MER2 repeat: matches 9. .94 of consensus"  
 8552. 8841  
 /note="AluSx repeat: matches 1. .292 of consensus"  
 9256. 9678  
 /gene="dJ322G13.1"  
 /note="match: proteins: Tr:Q9UKK6 Tr:Q9QZV9 Tr:Q9Y3M5  
 Tr:Q9V3H8 Tr:Q9Y3M4 Tr:Q9U757 Wp:V71P9AM.B Sw:Q42242"  
 /codon\_start=1  
 /evidence=not\_experimental  
 /product="dJ322G13.1 (NMF2-related export protein NXT1)"  
 /protein\_id="CAC03437.2"  
 /db\_xref="GI:10862837"  
 /translation="MASVDKFTVYDQACRAAEFVNYVTMDKRRRLSLVYGAT  
 LVNKGNAVSGGESLSEPEFPPSEFQSYVDQPVHDEAFPSQTVLYVIGCSVKFE  
 GNMKRDNFQNFILTAQSPSTWTKVIAISDCRFQDMS"  
 complement(9828. .9960)  
 /note="match: STS: Em:G42938"  
 10134. 10209  
 /note="MER5B repeat: matches 3. .82 of consensus"  
 10219. 10270  
 /note="13 copies 4 mer ggtg 96% conserved"  
 10220. 10271  
 /note="26 copies 2 mer tg 98% conserved"  
 10223. 10278  
 /note="4 copies 14 mer 92% conserved"  
 10299. 10431  
 /note="AluSg/x repeat: matches 128. .259 of consensus"  
 10445. 10487  
 /note="MER5B repeat: matches 34. .76 of consensus"  
 11048. 11269  
 /note="LTR1J repeat: matches 125. .364 of consensus"  
 complement(11808. .13219)  
 /gene="dJ322G13.5"  
 complement(join(<11808. .12183,12265. .12351,13169. .>13219))  
 /gene="dJ322G13.5"  
 /product="dJ322G13.5.1 (putative novel transcript, isoform 1)"  
 /note="match: ESTs: Em:AI693606 Em:AI984374"  
 /evidence=not\_experimental  
 12286. 12878  
 /note="match: STS: Em:HS638D8S"  
 12555. 12780  
 /note="AluSg/x repeat: matches 80. .305 of consensus"  
 12988. 13655  
 /note="Cpg Island"  
 /evidence=not\_experimental  
 14120. 14849  
 /note="MER49 repeat: matches 184. .916 of consensus"  
 14936. 15089  
 /note="AluSx repeat: matches 138. .282 of consensus"

```

repeat_region 15090..15383
/Note="AluX repeat: matches 1..294 of consensus"
15384..15518
/Note="AluX repeat: matches 1..138 of consensus"
15519..15698
repeat_region 15699..15928
/Note="L2 repeat: matches 2709..2750 of consensus"
15714..16375
/Note="MER67C repeat: matches 1..710 of consensus"
16723..17983
/Note="CpG Island"
/evidence-not-experimental
18803..18837
repeat_region 18838..19228
/Note="MIR repeat: matches 126..160 of consensus"
19229..19287
misc_feature 19288..19288
/Note="match: GSS: Em: A0203877"
19289..19289
/Note="match: GSS: Em: A0145317"
18914..18973
repeat_region 18974..19249
/Note="MIR repeat: matches 74..140 of consensus"
19249..19878
misc_feature 19879..19879
/Note="match: GSS: Em: A0895221"
19598..19598
/Note="match: GSS: Em: A0322613.2"
19599..19599
/Note="match: GSS: Em: A0322613.2"
25305..28277
/gene="d3322613.2"
/product="d3322613.2.3 (zinc finger protein FLJ21794, isoform 3)"
/Note="match: CDNA: Em: AL110217 Em: AF167320 Em: X78927 Em: AJ224763 Em: AL080123 Em: AF161544 Em: Z67747 Em: AB037770 Em: X60155 Em: X82125 Em: U66561 Em: AF022146 Em: L36315 Em: M36514 Em: X12592 Em: X64413 Em: AF022158 Em: L75847 Em: D10632 Em: X06021 Em: X89264 Em: X78933 Em: L15309 Em: X52533 Em: S67970 Em: AK025447
match: ESTs: Em: AA17107 Em: R89950 Em: AT023363 Em: A1217180 Em: AM291034 Em: AA17012 Em: AA701588 Em: A1177089 Em: A1145986 Em: AW079910 Em: A1961159 Em: AW019926 Em: AA035213 Em: AW082641 Em: AW594742 Em: AW531679 Em: A1984655 Em: AA273573 Em: A173862 Em: AV236342 Em: AW144279 Em: AL135710 Em: AA365129 Em: AA346014 Em: AA062926 Em: AL046380 Em: AA254007 Em: T85404 Em: AA383356
Query Match 100.0% Score 830 DB 9 Length 112180
Best Local Similarity 100.0% Pred. No. 4.7e-225 Indels 0 Gaps 0
Matches 830 Conservative 0 Mismatches 0

```

```

OY 421 CAACAACGGGACTTACACAGAACTTATCTGACCGCCCGGCTCACCACCAAC 480
DB 9567 CAACAACGGGACTTACACAGAACTTATCTGACCGCCCGGCTCACCACCAAC 9526
OY 481 AGTGTGAAGATGCCAGTACCTCTCCGCTTCAGAGACTGGCCAGCTAGTGGGTG 540
DB 9627 AGTGTGAAGATGCCAGTACCTCTCCGCTTCAGAGACTGGCCAGCTAGTGGGTG 9686
OY 541 GCAGAGCTCTCTTGCCTTACCTTACGCTTACGCTTGTAGAGAAATGCAACCTGCATCT 600
DB 9687 GCAGAGCTCTCTTGCCTTACCTTACGCTTACGCTTGTAGAGAAATGCAACCTGCATCT 9746
OY 601 CAAGATGTAGGAACAACAACTTATCTGTTGTCGAGACATGACATCTCACTG 660
DB 9747 CAAGATGTAGGAACAACAACTTATCTGTTGTCGAGACATGACATCTCACTG 9806
OY 661 TGGCGAGTGAACCTCTTTTGTGTGTGTCAGAGTCTAGAGTCCCTTCTGATATATA 720
DB 9807 TGGCGAGTGAACCTCTTTTGTGTGTGTCAGAGTCTAGAGTCCCTTCTGATATATA 9866
OY 721 CTGTGTGATAGTATCTTCTTCAAGTATGTAACCTTCTATTTCTACTGCCCAG 780
DB 9867 CTGTGTGATAGTATCTTCTTCAAGTATGTAACCTTCTATTTCTACTGCCCAG 9926
OY 781 TAGAGACTGTGATCTGGAATCTGACAAATAATTAATATACATG 830
DB 9927 TAGAGACTGTGATCTGGAATCTGACAAATAATTAATATACATG 9976

RESULT 2
BC003410
LOCUS
DEFINITION
Homo sapiens, NFE2-related export protein 1, clone MGC:4978
IMAGE:3450767, mRNA, complete cds.
ACCESSION
BC003410
VERSION
BC003410.1 GI:13097317
KEYWORDS
MGC.
SOURCE
Homo sapiens.
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 1176)
AUTHORS
Strausberg, R.
TITLE
Direct Submission
JOURNAL
Submitted (20-FEB-2001). National Institutes of Health, Mammalian
Gene Collection (MGC). Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK
NIH-MGC Project URL: http://mgc.ncl.nih.gov
Contact: MGC help desk
Email: gcgaps@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: villalona@bcm.tmc.edu.
Villalona, D.K., Luna, R.A., Hale, S.M., Huliy, S., Lu, X., Garcia,
A.M., Hollway, M., Telford, B., Hodgson, A., Bouck, J., Yu, W.,
Muzny, D.M., Gibbs, R.A.
FEATURES
source
1..1176
/organism="Homo sapiens"
/db_xref="locusid:29107"

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>  
 Series: IRAC Plate: 3 Row: P Column: 18  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gl: 7019470.

```

/db_xref="taxon:9606"
/clone="MGC:4978 IMAGE:3450767"
/tissue_type="Placenta, choriochorionoma"
/clone_lib="NH_MGC_10"
/lab_host="DH10B"
/notes="vector: pcMV-sport6"
388..810
/codon_start=1
/product="NF2-related export protein 1"
/protein_id="AA03410.1"
/db_xref="GI:13097318"
/translation="MASVDEKTYVDQACRAAEFVNYYTMDKRRRLSLRYMGAT
LVMNGNAVSGESLSEFEMLPSSEFOISVDCOPVHDEATPSOTTVLVVIGSVKFE
GNKORDFNFIITAOASPSNTVMKIASDCFRFODMAS"

BASE COUNT      294 a      309 c      316 g      257 t
ORIGIN
Query Match      94.8%; Score 786.8; DB 9; Length 1176;
Best Local Similarity 99.7%; Pred. No. 6.5e-213;
Matches 788; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 41 CCTGCAACCCCTGTTCCCAAGCAGAGAAATACCTGTGGAGCCCTCTCCATA 100
DB 11 |||||
DB 319 CCGCGACCCCTGTTCCCAAGCAGAGAAATACCTGTGGAGCCCTCTCCATA 378
QY 101 GAACAGAGATGGCATGTGTGATTTCAAGACCTATGTGATCAGGCTGACAGCTGCT 160
DB 11 |||||
DB 379 GAACAGAGATGGCATGTGTGATTTCAAGACCTATGTGATCAGGCTGACAGCTGCT 438
QY 161 GAGGAGTTTGTCAATGTCTACTACACCAACATGATGAGCGGCGGCTTGTCTCCGC 220
DB 11 |||||
DB 439 GAGGAGTTTGTCAATGTCTACTACACCAACATGATGAGCGGCGGCTTGTCTCCGC 498
QY 221 CTGTACATGGGACACACCCCTGTCTGGAATGCAATGCTGTTTCAGAGACAATCC 280
DB 11 |||||
DB 499 CTGTACATGGGACACACCCCTGTCTGGAATGCAATGCTGTTTCAGAGACAATCC 558
QY 281 TTGAGTGAATTTTGAATGTGCTCCACGAGTTCACCAATCAGGCTGTAGATGAC 340
DB 11 |||||
DB 559 TTGAGTGAATTTTGAATGTGCTCCACGAGTTCACCAATCAGGCTGTAGATGAC 618
QY 341 CAGCCTGTTCATGATGAGGACCAACACCAACGACGAGCTGCTGTTGTCTGATGTA 400
DB 11 |||||
DB 619 CAGCCTGTTCATGATGAGGACCAACACCAACGACGAGCTGCTGTTGTCTGATGTA 678
QY 401 TCAGTGAATTTGAGGGGAAACAACAAGGAGTTCACCAACTTCTCTGACCGCC 460
DB 11 |||||
DB 679 TCAGTGAATTTGAGGGGAAACAACAAGGAGTTCACCAACTTCTCTGACCGCC 738
QY 461 CAGGCTCTCCAGACACACAGTGTGGAAGATGCAAGGAGTTCGCTCCGCTCAGAG 520
DB 11 |||||
DB 739 CAGGCTCTCCAGACACACAGTGTGGAAGATGCAAGGAGTTCGCTCCGCTCAGAG 798
QY 521 TGGGCGAGTGTGGGGGTGGGAGAGGTCTCTTGTGCTATTCAGGCTGAGCTGTAGA 580
DB 11 |||||
DB 799 TGGGCGAGTGTGGGGGTGGGAGAGGTCTCTTGTGCTATTCAGGCTGAGCTGTAGA 858
QY 581 GAAATGCAACCTGCACTCTCAAGATGTGAGGAACACAAAGTTCATTTCTGTTGGG 640
DB 11 |||||
DB 859 GAAATGCAACCTGCACTCTCAAGATGTGAGGAACACAAAGTTCATTTCTGTTGGG 918
QY 641 AAGACAGTGAAGTCCAGTGTGCGAGGTTGAACCTTTTGTGCTCAATTTCTAGGA 700
DB 11 |||||
DB 919 AAGACAGTGAAGTCCAGTGTGCGAGGTTGAACCTTTTGTGCTCAATTTCTAGGA 978
QY 701 GTCCCTTCTCAATATATATCTGTTGTGATGATTTCTTTTCAAGAGTAGTAATTT 760
DB 11 |||||
DB 979 GTCCCTTCTCAATATATATCTGTTGTGATGATTTCTTTTCAAGAGTAGTAATTT 1038
QY 761 CTAATTTTCTACTTGGCCAGTAGAGACTGTGATTCGGAATTTGACAAATTAATTT 820
DB 11 |||||
DB 1039 CTAATTTTCTACTTGGCCAGTAGAGACTGTGATTCGGAATTTGACAAATTAATTT 1098

```

```

QY 821 AATACACATG 830
DB 1099 AATACACATG 1108

RESULT 3
AF156957
LOCUS
DEFINITION
Homo sapiens NF2-related export protein Nxt1 (Nxt1) mRNA, complete
cde
ACCESSION
AF156957
AF156957.1 GI:580864
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
1 (bases 1 to 965)
Black, B.E., Levesque, L., Holsaka, J.M., Wood, T.C. and Paschal, B.M.
Identification of an NF2-related factor that binds Ran-GTP and
regulates nuclear protein export
Mol. Cell. Biol. 19 (12), 8616-8624 (1999)
JOURNAL
MEDLINE
20036817
PUBMED
10567585
REFERENCE
2 (bases 1 to 965)
Black, B.E. and Paschal, B.M.
Direct Submission
Submitted (09-JUN-1999) Center for Cell Signaling, University of
Virginia, 7161 Hospital West, Box 577 Health Sciences Center,
Charlottesville, VA 22908, USA
FEATURES
source
1..965
location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
1..965
/gene="NXT1"
272..694
/gene="NXT1"
/function="stimulates nuclear protein export"
/notes="similar to the sequence presented in Genbank
Accession Number AA85905; nuclear transport factor; binds
Ran-GTP"
/codon_start=1
/product="NF2-related export protein Nxt1"
/protein_id="AA054942.1"
/db_xref="GI:580865"
/translation="MASVDEKTYVDQACRAAEFVNYYTMDKRRRLSLRYMGAT
LVMNGNAVSGESLSEFEMLPSSEFOISVDCOPVHDEATPSOTTVLVVIGSVKFE
GNKORDFNFIITAOASPSNTVMKIASDCFRFODMAS"

BASE COUNT      207 a      265 c      263 g      230 t
ORIGIN
Query Match      91.5%; Score 759.8; DB 9; Length 965;
Best Local Similarity 99.7%; Pred. No. 3.1e-205;
Matches 761; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 41 CCTGCAACCCCTGTTCCCAAGCAGAGAAATACCTGTGGAGCCCTCTCCATA 100
DB 11 |||||
DB 203 CCGCGACCCCTGTTCCCAAGCAGAGAAATACCTGTGGAGCCCTCTCCATA 262
QY 101 GAACAGAGATGGCATGTGTGATTTCAAGACCTATGTGATCAGGCTGACAGCTGCT 160
DB 11 |||||
DB 263 GAACAGAGATGGCATGTGTGATTTCAAGACCTATGTGATCAGGCTGACAGCTGCT 322
QY 161 GAGGAGTTTGTCAATGTCTACTACACCAACATGATGAGCGGCGGCTTGTCTCCGC 220
DB 11 |||||
DB 323 GAGGAGTTTGTCAATGTCTACTACACCAACATGATGAGCGGCGGCTTGTCTCCGC 382
QY 221 CTGTACATGGGACACACCCCTGTGTGATGCAATGCAATGCTTTCAGAGACAATCC 280
DB 11 |||||
DB 383 CTGTACATGGGACACACCCCTGTGTGATGCAATGCAATGCTTTCAGAGACAATCC 442
QY 281 TTGAGTGAATTTTGAATGTGCTTCCAGCGAGTTCACCAATCAGCGTGTAGACTGC 340

```

```

Db 443 TTGAATGAGTTTGGAAATGTTGCTTCACAGGAGTTCCAAATCAGCGTGTAGACTGC 502
Oy 341 CAGCGCTTCATGATGAAAGCCACCAAGCCAGACAGGTCCTGTGTCTCTGTGA 400
Db 503 CAGCGCTTCATGATGAAAGCCACCAAGCCAGACAGGTCCTGTGTCTCTGTGA 562
Oy 401 TCAGTGAAGTTGAGGGGAGCAACAGCGGACTTCAACAGAACTTATCTGTACGCC 460
Db 563 TCAGTGAAGTTGAGGGGAGCAACAGCGGACTTCAACAGAACTTATCTGTACGCC 622
Oy 461 CAGCGCTTCACCCAGCAACAGAGTGTGAAGATCGCAAGTGTGCTTCCGCTTCCAGAGC 520
Db 623 CAGCGCTTCACCCAGCAACAGAGTGTGAAGATCGCAAGTGTGCTTCCGCTTCCAGAGC 682
Oy 521 TGGGCGACCTAGTGGGGGTGGCAGAGTCTCTTCTTCATTCACGCTAGCTGTAGA 580
Db 683 TGGGCGACCTAGTGGGGGTGGCAGAGTCTCTTCTTCATTCACGCTAGCTGTAGA 742
Oy 581 GAAATGCAAACTGCACTCAGAGATGTGAGAGACAGTTCATCTGTGTGGG 640
Db 743 GAAATGCAAACTGCACTCAGAGATGTGAGAGACAGTTCATCTGTGTGGG 802
Oy 641 AGACACTGCAAGCTCAGCTGTCGCGAGGTGAACTCTTTTGTGTCAAGTTCTAGA 700
Db 803 AGACACTGCAAGCTCAGCTGTCGCGAGGTGAACTCTTTTGTGTCAAGTTCTAGA 862
Oy 701 GTCCCTTTCCTGATATATATCTGTTGTCTAGTTCCTTCTTCAAGTAGTAACTTT 760
Db 863 GTCCCTTTCCTGATATATATCTGTTGTCTAGTTCCTTCTTCAAGTAGTAACTTT 922
Oy 761 CTATTTTCTACTGTCGCCAGTAGAGACTGATTTCTGAAAT 803
Db 923 CTATTTTCTACTGTCGCCAGTAGAGACTGATTTCTGAAAT 965

```

```

RESULT 4
BC002687 939 bp mRNA linear PRI 12-JUL-2001
DEFINITION Homo sapiens, NTF2-related export protein 1, clone MGC:3469
IMAGE:3608037, mRNA, complete cds.
ACCESSION BC002687
VERSION BC002687.1 GI:12803702
KEYWORDS MGC.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Strausberg, R.
Direct Submission
Submitted (05-FEB-2001) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA.
NIH-MGC Project URL: http://mgc.ncl.nih.gov
Contact: MGC help desk
Email: cgapbs-rt@mail.nih.gov
Tissue: Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: http://www.nisc.nih.gov/
nisc.mgc@nih.gov
Contact:
Shevchenko, Y., Wetherby, R.D., Beckstrom-Sternberg, S.M.,
Benjamin, B., Blakesley, R.W., Bouffard, G.G., Brinkley, C., Brooks, S.,
Dietrich, N.L., Guan, X., Gupta, J., Ho, S.-L., Karlins, E., Legaspi, R.,
Lim, M., Maduro, Q.L., Mastello, C., Mastrian, S.D., McCloskey, J.C.,
McMowell, J., Pearson, R., Snyder, B., Stantirpop, S., Thomas, P.J.,
Tjongson, E.E., Touchman, J.W., Tsurgent, C., Vogt, J.L., Walker, M.A.,
Zhang, L.-H. and Green, E.D.

```

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>  
 Series: IRAL Plate: 12 Row: m Column: 11  
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 7019470.

FEATURES  
 source  
 1. 939  
 /organism="Homo sapiens"  
 /db\_xref="LocusID:29107"  
 /db\_xref="taxon:9606"  
 /clone="MGC:3469 IMAGE:3608037"  
 /tissue\_type="uterus, endometrium adenocarcinoma"  
 /clone\_id="NIH\_MGC\_44"  
 /lab\_host="DH10B-R"  
 /note="vector: pORF7"  
 252..674  
 /codon\_start=1  
 /product="NTF2-related export protein 1"  
 /protein\_id="AA02687.1"  
 /db\_xref="gi:12803703"  
 /translation="MAYDFKTYVDQARAAEFVNVYTTTMDKRRRLSLYMGAT  
 LVNMGNAVSGOESLSEFPEMLPSSEFQISVDCCPVHDEATPSQTTVLVIGSVKFE  
 GNKORDFNFIITROASPSMTVMKIASDCRFQDMAS"

BASE COUNT 217 a 253 c 250 g 219 t

ORIGIN

Query Match 88.7%; Score 736.4; DB 9; Length 939;  
 Best Local Similarity 99.2%; Pred. No. 1.5e-198;  
 Matches 740; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

Oy 41 CCTCGAGCCCTGCTTCCCAAGCAGAGAAATACCTGTGGAGCCCTCTCCATA 100
Db 183 CCGCCAGCCCTGCTTCCCAAGCAGAGAAATACCTGTGGAGCCCTCTCCATA 242
Oy 101 GAACAGAGATGAGTCTGTGATTTCAAGACTTGTGATCAGAGCCCTGAGACTGCT 160
Db 243 GAACAGAGATGAGTCTGTGATTTCAAGACTTGTGATCAGAGCCCTGAGACTGCT 302
Oy 161 GAGGAGTTTGTCAAGTCTACTACACCAACATGATTAAGCGCGCGTTGTGTCGCCG 220
Db 303 GAGGAGTTTGTCAAGTCTACTACACCAACATGATTAAGCGCGCGTTGTGTCGCCG 362
Oy 221 CTGTACATGAGGACAGCCACCTGCTGTGAATGGCAATGCTGTTCAAGACAGAATCC 280
Db 363 CTGTACATGAGGACAGCCACCTGCTGTGAATGGCAATGCTGTTCAAGACAGAATCC 422
Oy 281 TTGATGAGTTTGTGAATGTGCTTCCAGCGAGTTCCAAATACGCTGTAGACTGC 340
Db 423 TTGATGAGTTTGTGAATGTGCTTCCAGCGAGTTCCAAATACGCTGTAGACTGC 482
Oy 341 CAGCGCTTCATGATGAAAGCCACCAAGCCAGACAGGTCCTGTGTCTATCTGTGA 400
Db 483 CAGCGCTTCATGATGAAAGCCACCAAGCCAGACAGGTCCTGTGTCTATCTGTGA 542
Oy 401 TCAGTGAAGTTGAGGGGAGCAACAGCGGACTTCAACAGAACTTATCTGTACGCC 460
Db 543 TCAGTGAAGTTGAGGGGAGCAACAGCGGACTTCAACAGAACTTATCTGTACGCC 602
Oy 461 CAGCGCTTCACCCAGCAACAGAGTGTGAAGATCGCAAGTGTGCTTCCGCTTCCAGAGC 520
Db 603 CAGCGCTTCACCCAGCAACAGAGTGTGAAGATCGCAAGTGTGCTTCCGCTTCCAGAGC 662
Oy 521 TGGGCGACCTAGTGGGGGTGGCAGAGTCTCTTCTTCATTCACGCTAGCTGTAGA 580
Db 663 TGGGCGACCTAGTGGGGGTGGCAGAGTCTCTTCTTCATTCACGCTAGCTGTAGA 722
Oy 581 GAAATGCAAACTGCACTCAGAGATGTGAGAGACAGTTCATTTCTGTGTGGG 640
Db 723 GAAATGCAAACTGCACTCAGAGATGTGAGAGACAGTTCATTTCTGTGTGGG 782
Oy 641 AGACACTGCAAGCTCAGCTGTCGCGAGGTGAACTCTTTTGTGTCTCAAGTTCTAGA 700

```



Db 783 AGACACTGACAGACTCCAGTGGCCGAGGTGAACTCTTTTGTGCTCAAGTCTAGGA 842  
Qy 701 GTCCCTTCTGTAATATATCTGTTGTCATATGTTCTTCCATCAAGTAACTTTT 760  
Db 843 GTCCCTTCTGTAATATATCTGTTGTCATATGTTCTTCCATCAAGTAACTTTT 902  
Qy 761 CTAATTTTCTACTGCGCAGTAGAGA 786  
Db 903 CTAATTTTCTACTGCGCAGTAGAGA 928

RESULT 5  
BC000759  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

BC000759 963 bp mRNA linear PRI 12-JUL-2001  
Homo sapiens, NTF2-related export protein 1, clone MGC:4329  
IMAGE:2820775, mRNA, complete cds.  
BC003029.1 GI:12804338  
MGC.  
Homo sapiens.  
Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Strausberg, R.  
Direct Submission  
Submitted (06-FEB-2001) National Institutes of Health, Mammalian  
Gene Collection (MGC), Cancer Genomics Office, National Cancer  
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,  
USA  
NIH-MGC Project URL: <http://mgc.nci.nih.gov>  
Contact: MGC help desk  
Email: [cgabs-remail.nih.gov](mailto:cgabs-remail.nih.gov)  
Tissue Procurement: DCTD/OTF  
CDNA Library Preparation: Rubin Laboratory  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Genome Sequence Centre,  
BC Cancer Agency, Vancouver, BC, Canada  
[info@cgsc.bc.ca](mailto:info@cgsc.bc.ca)  
Steven Jones, Jennifer Asano, Ian Bosdet, Yaron Butterfield,  
Susanna Chan, Rehdan Chlu, Chris Fjell, Erin Garland, Ran Guln,  
Leticia Hsiao, Martin Kirzysinski, Reta Kutsche, Oliver Lee, Soo  
Sen Lee, Victor Ling, Carrie Mathewson, Candice McLeavy, Steven  
Ness, Pawan Pandoh, Anna-Liisa Prahnu, Parvaneh Saeedi, Jacqueline  
Schell, Duane Smalius, Michael Smith, Lorraine Spence, Jeff Scott,  
Michael Thorne, Miranda Tsai, Natasja van den Bosch, Jill Vardy,  
George Yang, Scott Zuyderduyn, Marco Marra.

FEATURES  
Source  
Location/Qualifiers  
1. 963  
/organism="Homo sapiens"  
/db\_xref="locusid:29107"  
/db\_xref="taxon:9606"  
/clone="MGC:4329 IMAGE:2820775"  
/tissue\_type="Lung, small cell carcinoma"  
/clone\_id="NIH\_MGC\_7"  
/lab\_host="DH10B-R"  
/note="Vector: pOTB7"  
276..698  
/codon\_start=1  
/product="NTF2-related export protein 1"  
/protein\_id="AAH03029.1"  
/db\_xref="GI:12804339"  
/translation="MAVDKTYVDACRAAEFVNYVTMCKRRLLSRVWGTA  
LVNMGNAVSGDESISEFEKMPSEFISVYDCQPVYDEATPSSQTVLVVYICGVKFE  
GNQORDNQNFILTAQSPSNYKIASDFRQDMS"

BASE COUNT 220 a 263 c 256 g 224 t

ORIGIN

Query Match 88.3%; Score 733.2; DB 9; Length 963;  
Best Local Similarity 98.9%; Pred. No. 1,2e-197;  
Matches 738; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 41 CCCTGACGCCCTGTTGCCCAAGCAGAGAAATACCTGTGTGAGACCTCTCCATA 100  
Db 207 CCGCGCAGGCCCTGTGTTCCCAAGCAGAGAAATACCTGTGTGAGACCTCTCCATA 266  
Qy 101 GAAACAGAGATGGCATCTGTGATTTCAAGCTATGTGATAGAGCTCAGAGCTCT 160  
Db 267 GAACAGAGATGGCATCTGTGATTTCAAGCTATGTGATAGAGCTCAGAGCTCT 326  
Qy 161 GAGAGATTTGTCAATGTCTACTACACCAACCATGATTAAGGGGGGCTTGTCTCCGC 220  
Db 327 GAGAGATTTGTCAATGTCTACTACACCAACCATGATTAAGGGGGGCTTGTCTCCGC 386  
Qy 221 CTGTACATGGGCGACAGCCACCTGTGTGATGGAATGCGATCTGTTTCAAGACAAGATCC 280  
Db 387 CTGTACATGGGCGACAGCCACCTGTGTGATGGAATGCGATCTGTTTCAAGACAAGATCC 446  
Qy 281 TTGAGTGAAGTTTGTGAATGTGCTTCCAGGAGTTCCAAATACAGCTGTGAGACTGC 340  
Db 447 TTGAGTGAAGTTTGTGAATGTGCTTCCAGGAGTTCCAAATACAGCTGTGAGACTGC 506  
Qy 341 CAGCCTGTTGATGATGAGCCACACAGCCAGACAGGCTGTTGTTCATCTGTGA 400  
Db 507 CAGCCTGTTGATGATGAGCCACACAGCCAGACAGGCTGTTGTTCATCTGTGA 566  
Qy 401 TCAGTGAAGTTTGAAGGGAACAACAAGGAGCTTCAACAGAACTTCACTGACCGCC 460  
Db 567 TCAGTGAAGTTTGAAGGGAACAACAAGGAGCTTCAACAGAACTTCACTGACCGCC 626  
Qy 461 CAGGCTCACCAGCAACAGTGTGGAATGCAAGTACAGTCTCGCTCCAGAGAC 520  
Db 627 CAGGCTCACCAGCAACAGTGTGGAATGCAAGTACAGTCTCGCTCCAGAGAC 686  
Qy 521 TGGGCGAGTAAAGTGGGGGTGGCAGAGTCTCTTGTTCATTCAGCCCTGACTGTAGA 580  
Db 687 TGGGCGAGTAAAGTGGGGGTGGCAGAGTCTCTTGTTCATTCAGCCCTGACTGTAGA 746  
Qy 581 GAAATGCAACCTGACTCTCAAGATGTGAGAGACCAAGTATCTTGTGTTCGG 640  
Db 747 GAAATGCAACCTGACTCTCAAGATGTGAGAGACCAAGTATCTTGTGTTCGG 806  
Qy 641 AGACACTGACAGCTCCACTGTGCGAGGTGAATCTTTTGTGTGCAAGTCTAGGA 700  
Db 807 AGACACTGACAGCTCCACTGTGCGAGGTGAATCTTTTGTGTGCAAGTCTAGGA 866  
Qy 701 GTCCCTTCTGTAATATATCTGTTGTCATATGTTCTTCCATCAAGTAACTTTT 760  
Db 867 GTCCCTTCTGTAATATATCTGTTGTCATATGTTCTTCCATCAAGTAACTTTT 926  
Qy 761 CTAATTTTCTACTGCGCAGTAGAGA 786  
Db 927 CTAATTTTCTACTGCGCAGTAGAGA 952

RESULT 6  
BC000759  
LOCUS  
DEFINITION  
ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
AUTHORS  
TITLE

BC000759 972 bp mRNA linear PRI 12-JUL-2001  
Homo sapiens, NTF2-related export protein 1, clone MGC:2701  
IMAGE:2820775, mRNA, complete cds.  
BC000759.1 GI:12653930  
MGC.  
Homo sapiens.  
Homo sapiens.  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
Strausberg, R.  
Direct Submission



ORIGIN

Query Match: 86.7%; Score 719.2; DB 9; Length 1049;  
Best Local Similarity 98.7%; Pred. No. 1.2e-193;  
Matches 726; Conservative 0; Mismatches 8; Indels 2; Gaps 1;

QY 41 CCTCGAGCCCTGGTCCCAAGGAGGAAATACCTGGTGAAGCCCTCTTCATA 100  
DB 299 CCGGAGAGCCCTGGTCCCAAGGAGGAAATACCTGGTGAAGCCCTCTTCATA 358  
QY 101 GAACGAGATGACCTGTGGATTCAAGACTATGATGATGATGATGATGATGAT 160  
DB 359 GAACGAGATGACCTGTGGATTCAAGACTATGATGATGATGATGATGATGAT 418  
QY 161 GAGAGATTGTGAATGTCTACTACACACCATGATGATGATGATGATGATGATGAT 220  
DB 419 GAGAGATTGTGAATGTCTACTACACACCATGATGATGATGATGATGATGATGAT 478  
QY 221 CTGTACATGGGACACACACCTGTGTGGAATGGCAATGCTGTTCAAGACAAATCC 280  
DB 479 CTGTACATGGGACACACACCTGTGTGGAATGGCAATGCTGTTCAAGACAAATCC 538  
QY 281 TTGAGTGAATTTTGAATTTGCTTCCAGAGATGATGATGATGATGATGATGATGAT 340  
DB 539 TTGAGTGAATTTTGAATTTGCTTCCAGAGATGATGATGATGATGATGATGATGAT 598  
QY 341 CAGCCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 400  
DB 599 CAGCCTGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 658  
QY 401 TCAGTGAATTTGAGGGGAAACAAACAGGGGATTCACAGAACTTCACTGACCGCC 460  
DB 659 TCAGTGAATTTGAGGGGAAACAAACAGGGGATTCACAGAACTTCACTGACCGCC 718  
QY 461 CAGGCTCACCAGACACACAGTGTGAGATGCAAGTGTGCTGCTGCTGCTGCTGCTG 520  
DB 719 CAGGCTCACCAGACACACAGTGTGAGATGCAAGTGTGCTGCTGCTGCTGCTGCTG 778  
QY 521 TGGGCGAGCTAGTGGGGGTGGGAGAGTCTCTTCTTCACTGAGCCCTGCTGTAGA 580  
DB 779 TGGGCGAGCTAGTGGGGGTGGGAGAGTCTCTTCTTCACTGAGCCCTGCTGTAGA 836  
QY 581 GAATGCAAACTCTGACTCTCAAGATGTGAGGAAACAGTTCATTCTGTTGCGG 640  
DB 837 GAATGCAAACTCTGACTCTCAAGATGTGAGGAAACAGTTCATTCTGTTGCGG 696  
QY 641 AGACATGAGACACTGCTGCGGAGGTGAACCTTTTGTGCTCAAGTCTAGGA 700  
DB 897 AGACATGAGACACTGCTGCGGAGGTGAACCTTTTGTGCTCAAGTCTAGGA 956  
QY 701 GTCCCTTCTGTAATATATCTGTTGATGATGATGATGATGATGATGATGATGAT 760  
DB 957 GTCCCTTCTGTAATATATCTGTTGATGATGATGATGATGATGATGATGATGAT 1016  
QY 761 CTATTTTCTACTGCGCAGTAGA 786  
DB 1017 CTATTTTCTACTGCGCAGTAGA 1042

TITLE  
JOURNAL  
COMMENT

Direct Submission  
Submitted (24-FEB-2001) Sanger Centre, Hinxton, Cambridgeshire,  
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
Requests: clonerequest@sanger.ac.uk  
On Feb 28, 2001 this sequence version replaced g1:12831820:  
During sequence assembly data is compared from overlapping clones.  
Where differences are found these are annotated as variations  
together with a note of the overlapping clone name. Note that the  
variation annotation may not be found in the sequence submission  
corresponding to the overlapping clone, as we submit sequences with  
only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all  
regions were either double-stranded or sequenced with an alternate  
chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such  
as compressions and repeats; all regions were covered by at least  
one plasmid subclone or more than one M13 subclone; and the  
assembly was confirmed by restriction digest. The following  
abbreviations are used to associate primary accession numbers given  
in the feature table with their source databases: Em., EMBL; Sw.,  
SWISSPROT; Tr., TrEMBL; Wp., WORMPEP; information on the WORMPEP  
database can be found at  
[http://www.sanger.ac.uk/Projects/C\\_elegans/wormpep](http://www.sanger.ac.uk/Projects/C_elegans/wormpep) This sequence  
was generated from part of bacterial clone contigs of human  
chromosome 13, constructed by the Sanger Centre Chromosome 13  
Mapping Group. Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chr13>  
RP11-527F15 is from the library RP11-11.2 constructed by the group  
of Pieter de Jong. For further details see  
<http://www.chori.org/bacpac/home.htm>  
VECTOR: pBACe3.6

FEATURES

IMPORTANT: This sequence is not the entire insert of clone  
RP11-527F15. It may be shorter because we sequence overlapping  
sections only once, except for a 100 base overlap.  
The true left end of clone RP11-11F1 is at 135520 in this  
sequence. The true right end of clone RP11-186J16 is at 100 in this  
sequence.

source

Location/Qualifiers  
1..135619  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/clone="RP11-527F15"  
/clone\_1fb="RP11-11.2"  
55..179  
/note="MER5B repeat: matches 42..174 of consensus"  
396..567  
/note="FAM repeat: matches 1..168 of consensus"  
1056..1399  
/note="MER7A repeat: matches 1..346 of consensus"  
1805..2527  
/note="L1PA16 repeat: matches 5418..6157 of consensus"  
2699..3183  
/note="MER74A repeat: matches 5..528 of consensus"  
3218..3373  
/note="L2 repeat: matches 2289..2474 of consensus"  
4524..4982  
/note="MTRIC repeat: matches 1..466 of consensus"  
5006..5377  
/note="MTR1A2 repeat: matches 1..374 of consensus"  
5614..6265  
/note="LIM4 repeat: matches 3331..3985 of consensus"  
complement(6016..6421)  
/note="match: SRS: Em:G53026"  
6368..6618  
/note="MTRIF repeat: matches 1..255 of consensus"  
6619..7146  
/note="MER1A repeat: matches 2..527 of consensus"  
7147..7238  
/note="MTRIF repeat: matches 255..347 of consensus"  
7239..7538  
/note="AluY repeat: matches 1..302 of consensus"  
7539..7674

RESULT 8  
AL354809 135619 bp DNA linear PRI 25-FEB-2001  
LOCUS Human DNA sequence from clone RP11-527F15 on chromosome 13,  
DEFINITION complete sequence.  
ACCESSION AL354809  
VERSION AL354809.12 GI:13161606  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 135619)  
AUTHORS Tromans,A.

```
/note="MLTIF repeat: matches 347. .472 of consensus"
repeat_region 11850. 12435
/note="MER21B repeat: matches 100. .660 of consensus"
repeat_region 12624. 12694
/note="MER21B repeat: matches 716. .787 of consensus"
repeat_region 12948. 13260
/note="L1M8 repeat: matches 5623. .5933 of consensus"
repeat_region 14557. 14848
/note="L1R16A repeat: matches 7. .411 of consensus"
repeat_region 16437. 16488
/note="L2 repeat: matches 2643. .2694 of consensus"
repeat_region 17820. 18331
/note="MLTIF repeat: matches 1. .541 of consensus"
repeat_region 18332. 18430
/note="L2 repeat: matches 2641. .2744 of consensus"
repeat_region 19075. 19945
/note="L1M4 repeat: matches 4437. .5313 of consensus"
repeat_region 20939. 21003
/note="MIR repeat: matches 84. .144 of consensus"
repeat_region 22102. 22624
/note="L1M3 repeat: matches 7189. .7739 of consensus"
repeat_region 22625. 22799
/note="AluJ repeat: matches 122. .303 of consensus"
repeat_region 22834. 22968
/note="AluJ repeat: matches 1. .133 of consensus"
repeat_region 22973. 23766
/note="L1M3 repeat: matches 6414. .7196 of consensus"
repeat_region 23771. 24574
/note="L1M3 repeat: matches 5433. .6288 of consensus"
repeat_region 24855. 24958
/note="MIR repeat: matches 112. .215 of consensus"
repeat_region 26923. 26986
/note="32 copies 2 mer ag 90% conserved"
repeat_region 27043. 27129
/note="L2 repeat: matches 2432. .2520 of consensus"
repeat_region 27153. 27242
/note="L2 repeat: matches 2620. .2710 of consensus"
repeat_region 28309. 28610
/note="AluSg repeat: matches 1. .303 of consensus"
repeat_region 31533. 31624
/note="L1P3 repeat: matches 5. .96 of consensus"
repeat_region 31623. 34630
/note="L1P repeat: matches 847. .3842 of consensus"
repeat_region 34722. 34930
/note="L2 repeat: matches 2192. .2404 of consensus"
repeat_region 36155. 36927
/note="MER21B repeat: matches 8. .790 of consensus"
repeat_region 37387. 37909
/note="L2 repeat: matches 2158. .2708 of consensus"
repeat_region 39234. 39522
/note="MLT2E repeat: matches 118. .397 of consensus"
repeat_region 39853. 40305
/note="MLT2B repeat: matches 1. .444 of consensus"
repeat_region 41902. 41949
/note="24 copies 2 mer aa 72% conserved"
repeat_region 42011. 42058
/note="MLT2F repeat: matches 367. .414 of consensus"
repeat_region 42059. 42086
/note="7 copies 4 mer ggtt 100% conserved"
repeat_region 42089. 42452
/note="MLT2F repeat: matches 8. .374 of consensus"
repeat_region 42831. 43032
/note="L1P3 repeat: matches 6. .218 of consensus"
repeat_region 43033. 43256
/note="AluJ repeat: matches 1. .225 of consensus"
repeat_region 43257. 43521
/note="L1P3 repeat: matches 218. .446 of consensus"
repeat_region 43525. 43681
/note="AluJ repeat: matches 147. .311 of consensus"
repeat_region 43683. 43892
/note="L1P3 repeat: matches 530. .759 of consensus"
repeat_region 44049. 44120
/note="AluJ repeat: matches 223. .296 of consensus"

repeat_region 44121. 44271
/note="L1P3 repeat: matches 2552. .2708 of consensus"
repeat_region 45039. 45252
/note="L1P8 repeat: matches 5950. .6163 of consensus"
repeat_region 45253. 45309
/note="L1P3 repeat: matches 5681. .5737 of consensus"
repeat_region 45316. 45444
/note="L2 repeat: matches 2334. .2462 of consensus"
repeat_region 46424. 46491
/note="L2 repeat: matches 2581. .2652 of consensus"
repeat_region 46534. 46620
/note="L2 repeat: matches 2386. .2677 of consensus"
repeat_region 47570. 47684
/note="L1P3 repeat: matches 1. .117 of consensus"
repeat_region 47698. 50321
/note="L1P3 repeat: matches 3529. .6165 of consensus"
repeat_region 50339. 50386
/note="L1P3 repeat: matches 2 mer aa 83% conserved"
repeat_region 50389. 50659
/note="AluJ repeat: matches 1. .282 of consensus"
repeat_region 50660. 50777
/note="59 copies 2 mer ta 75% conserved"
repeat_region 50662. 50777
/note="29 copies 4 mer tata 75% conserved"
repeat_region 51545. 51618
/note="L1M3 repeat: matches 5883. .5952 of consensus"
repeat_region 51619. 51979
/note="L1P3 repeat: matches 1. .364 of consensus"
repeat_region 51980. 52081
/note="L1M3 repeat: matches 5804. .5883 of consensus"
repeat_region 52121. 52465
/note="L1M3 repeat: matches 5374. .5715 of consensus"
repeat_region 52466. 52624
/note="FAM repeat: matches 1. .165 of consensus"
repeat_region 52625. 52945
/note="L1M3 repeat: matches 5060. .5374 of consensus"
repeat_region 53108. 53409
/note="AluSg repeat: matches 1. .301 of consensus"
repeat_region 53907. 54022
/note="L1M3 repeat: matches 5305. .5433 of consensus"
repeat_region 55475. 55613
/note="MIR repeat: matches 3. .146 of consensus"

Query Match 61.3% Score 508.4; DB 9; Length 135619;
Best Local Similarity 95.8%; Pred. No. 2, 1e-133;
Matches 344; Conservative 0; Mismatches 21; Indels 3; Gaps 2;
```

```

Db      102385 CCCGACCTCTGTAGAGAAATGCAATCTGACCTCAGATGTGAGAACACAGTCA 102444
Oy      626 TTTCGTGTGTGGGAGACACGACACCTGCTGCCGAGCTGCACTC--TTTTTG 683
Db      102445 TTTCGTGTGTGTGGGAGACACGACACCTGCTGCCGAGCTGCACTC--TTTTTG 102504
Oy      684 TTGCTCAAGTCTAGAGAGCCCTTCCGTAATATATCTGTTGTCATAGTTCCTTT 743
Db      102505 TTGCTCAAGTCTAGAGAGCCCTTCCGTAATATATCTGTTGTCATAGTTCCTTT 102564
Oy      744 CAAAGTAGTAACCTTTCTATTCTCTA 771
Db      102565 CAAAGTAGTAACCTTTCTATTCTCTA 102592

RESULT 9
LOCUS   AL513495
DEFINITION Homo sapiens chromosome 13 clone RP11-466M19, *** SEQUENCING IN
ACCESSION AL513495
VERSION   AL513495.1 GI:12733863
KEYWORDS  HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE    human.
ORGANISM  Homo sapiens
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 147310)
AUTHORS   Burton,J.
TITLE     Direct Submission
JOURNAL   Submitted (12-JUN-2001) Sanger Centre, Hinxton, Cambridgeshire,
           CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk
           requests: clonerequests@sanger.ac.uk
           ----- Genome Center
           Center: Sanger Centre
           Center code: SC
           Web site: http://www.sanger.ac.uk
           Contact: humquerry@sanger.ac.uk
           ----- Project Information
           Center project name: ba466M19
           ----- Summary Statistics
           Sequencing vector: pGAP4; version 4.5
           Chemistry: Dye-terminator Big Dye; 100% of reads
           Consensus quality: 138138 bases at least Q40
           Consensus quality: 141205 bases at least Q30
           Consensus quality: 143097 bases at least Q20
           Insert size: 145810; sum-of-contrigs
           Insert size: 162901; agarose-fp
           Quality coverage: 3.13x in Q20 bases; sum-of-contrigs Quality
           coverage: 2.95x in Q20 bases; agarose-fp
           -----
           * NOTE: This is a 'working draft' sequence. It currently
           * consists of 16 contrigs. The true order of the pieces
           * is not known and their order in this sequence record is
           * arbitrary. Gaps between the contrigs are represented as
           * runs of N, but the exact sizes of the gaps are unknown.
           * This record will be updated with the finished sequence
           * as soon as it is available and the accession number will
           * be preserved.
           *
           * 1
           * 3020 3119: contrig of 3019 bp in length
           * 3120 12625: contrig of 9506 bp in length
           * 12626 12725: gap of 100 bp
           * 12726 20101: contrig of 7376 bp in length
           * 20102 20201: gap of 100 bp
           * 20202 23460: contrig of 3259 bp in length
           * 23461 23560: gap of 100 bp
           * 23561 34856: contrig of 11296 bp in length
           * 34857 34956: gap of 100 bp
           * 34957 50993: contrig of 16037 bp in length
           * 50994 51093: gap of 100 bp
           * 51094 62025: contrig of 10932 bp in length

```

```

FEATURES
source
1..147310
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="13"
/clone="RP11-466M19"
/clone_1b="RPC1-11.2"
1..3019
/note="assembly-fragment:00352
fragment_chain:1"
3120..12625
/note="assembly-fragment:01013
fragment_chain:1"
12726..20101
/note="assembly-fragment:00687
fragment_chain:1"
20202..23460
/note="assembly-fragment:00760
fragment_chain:1"
23561..34856
/note="assembly-fragment:00179
fragment_chain:1"
34957..50993
/note="assembly-fragment:00160
fragment_chain:2"
51094..62025
/note="assembly-fragment:00712
fragment_chain:2"
62126..65299
/note="assembly-fragment:00413
fragment_chain:3"
65400..74814
/note="assembly-fragment:00763
fragment_chain:3"
74915..78856
/note="assembly-fragment:00035"
78957..106521
/note="assembly-fragment:00290"
106622..119811
/note="assembly-fragment:00296"
119912..124426
/note="assembly-fragment:00410"
124527..137019
/note="assembly-fragment:00478"
137120..141832
/note="assembly-fragment:00564"
141933..147310
/note="assembly-fragment:00307
vector_side:right"
vector_end:SP6

BASE COUNT 44700 a 28018 c 28143 g 44941 t 1508 others

Query Match
Best Local Similarity 95.8%; Pred. No. 2.1e-133;
61.3%; Score 508.4; DB 2; Length 147310;

```

Matches 544; Conservative 0; Mismatches 21; Indels 3; Gaps 2;

207 GTTGTGCTCCCGCTGACATGGGACACACCCCTGCTGGAATGCAATGCTGTT 266  
 51966 GTGCACTGCGCCGCTGACAGGGGACAGCCACCTGCTGGAAGGCAATGCTGTT 52025  
 267 CAGGACAAAGATCCTTGTAGTGGTTTGTGAATGTGCTTCACGAGGTTCCAAATGA 326  
 52026 CAGGACAAAGATCCTTGTAGTGGTTTGTGAATGTGCTTCACGAGGTTCCAAATGA 52085  
 327 GGTGTGAGACAGCAGCCTGTCATGATGAAGCCACACAGCCAGCAGCTGCTTG 386  
 52086 GGTGTGAGACAGCAGCCTGTCATGATGAAGCCACACAGCCAGCAGCTGCTTG 52145  
 387 TTGTATCTGTGTGATCAGTGAAGTTTG -AGGGGAAACAAACAGGAGCTTCAACCGAAG 445  
 52146 TTGTATCTGTGTGATCAGTGAAGTTTG -AGGGGAAACAAACAGGAGCTTCAACCGAAG 52205  
 446 TTGATCTGTGACCGCCAGCCTGTCATGATGAAGCCACACAGCCAGCAGCTGCTTG 505  
 52206 TTGATCTGTGACCGCCAGCCTGTCATGATGAAGCCACACAGCCAGCAGCTGCTTG 52265  
 506 TTCCGCTTCCAGAGACTGGGCGCAGTATGAGGGGCTGCGAGAGTCTTCTTCATTCAG 565  
 52266 TTCCGCTTCCAGAGACTGGGCGCAGTATGAGGGGCTGCGAGAGTCTTCTTCATTCAG 52325  
 566 CCGTACCTGTGAGAGAAATGCAACCTGACTCTCAAGAGTGTGAGAACACAACTTCA 625  
 52326 CCGTACCTGTGAGAGAAATGCAACCTGACTCTCAAGAGTGTGAGAACACAACTTCA 52385  
 626 TTTCGTTGTCGCGAGACAGTACAGACTGCTGCGAGGTTGACTGTC -TTTTTG 683  
 52386 TTTCGTTGTCGCGAGACAGTACAGACTGCTGCGAGGTTGACTGTC -TTTTTG 52445  
 684 TTTCGTTGTCGCGAGACAGTACAGACTGCTGCGAGGTTGACTGTC -TTTTTG 743  
 52446 TTTCGTTGTCGCGAGACAGTACAGACTGCTGCGAGGTTGACTGTC -TTTTTG 52505  
 744 CAAAGTAGTAACCTTTCTATTTTCTA 771  
 52506 CAAAGTAGTAACCTTTCTATTTTCTA 52533

RESULT 10  
 AC110699/c 186415 bp DNA linear HTG 17-JUL-2002  
 LOCUS Rattus norvegicus clone CH230-157L12, \*\*\* SEQUENCING IN PROGRESS  
 DEFINITION \*\*\* 54 unordered pieces.  
 AC110699  
 AC110699.3 GI:21744668  
 HTG: HTGS\_PHASL1  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sclurognathu; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 186415)  
 Muzny,D.M., Adams,C., Adio-Oduola,B., All-ouman,F.R., Allen,C.,  
 Alsdbrook,S.L., Anaratunga,H.C., Are,J.R., Ayele,M., Banks,T.,  
 Barbarella,J., Benton,J., Bimaye,K., Blankenburg,K., Bonnin,D.,  
 Bouck,J., Bowles,S., Brileva,M., Brown,E., Brown,M., Bryant,N.P.,  
 Buhey,C., Burch,P., Burkett,C., Burrell,K.L., Byrd,N.C.,  
 Caron,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D.,  
 Chen,G., Chen,R., Chen,Z., Chowdhry,I., Christopoulos,C.,  
 Cleveland,C.D., Cox,C., Coyle,M.D., Dathorne,S.R., David,R.,  
 Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinu,H.H.,  
 Douthett,K.J., Draper,H., Duan-Rocha,S., Durbin,K.J.,  
 Earnhart,C., Edgar,D., Edwards,C.C., Elhaj,C., Escotto,M.,  
 Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P.,  
 Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R.,  
 Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K.,  
 Harris,C., Harris,K., Hart,M., Havlak,P., Hawes,A., Hernandez,J.,

Hernandez,O., Hodgeson,A., Hognes,M., Holloway,C., Hollins,B.,  
 Homel,F., Howard,S., Huber,J., Hulyk,S., Hume,J., Jackson,L.E.,  
 Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S.,  
 Karlsson,E., Kelly,S., Khan,U., Kling,L., Korvan,J., Kovar,C.,  
 Kratochvil,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,  
 Li,J., Li,Z., Lichtarge,O., Liew,C., Liu,J., Liu,W., Louised,H.,  
 Lozano,R.J., Lu,X., Lucier,A., Lucier,R., Luna,R., Ma,J.,  
 Maheshwari,M., Mapua,P., Martin,R., Martindale,A., Martinez,E.,  
 Massey,E., McLeod,M.P., Meador,M., Mel,G., Metzger,M.,  
 Miner,G., Miner,Z., Mitchell,T., Mohabat,K., Morgan,M., Morris,S.,  
 Moser,M., Neal,D., Newton,J., Newton,N., Nguyen,A., Nguyen,N.,  
 Nguyen,N., Nickerson,E., Nwokenkwo,S., Oguh,M., Okunolu,G.,  
 Oradunye,N., Oyler,D., Pace,A., Payton,B., Peery,J., Perez,L.,  
 Peters,L., Pickens,R., Primus,E., Pu,L., Quiles,M., Ren,Y.,  
 Rives,M., Rojas,A., Rojokan,I., Roite,M., Ruiz,S., Saverly,G.,  
 Scherer,S., Scott,G., Shen,H., Shoshani,N., Sisson,I.,  
 Sodergren,E., Somaite,T., Sparks,A., Stanley,H., Stone,H.,  
 Sutton,A., Svatek,A., Taber,P., Tamerisa,A., Tamerisa,K., Tang,H.,  
 Tansey,J., Taylor,C., Taylor,T., Telford,B., Thomas,N., Thomas,S.,  
 Usmani,K., Vasquez,L., Vera,V., Villalón,D., Vinson,R., Wang,Q.,  
 Wang,S., Ward-More,S., Warren,R., Washington,C., Wallington,S.,  
 Williams,G., Williamson,A., Wlezyk,R., Wooden,S., Worley,K.,  
 Wu,C., Wu,Y., Wu,Y.F., Zhou,J., Zorrilla,S., Nelson,D.,  
 Weinstock,G., and Gibbs,R.  
 Direct Submission  
 Unpublished  
 2 (bases 1 to 186415)  
 Worley,K.C.  
 Submitted (15-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 186415)  
 Worley,K.C.  
 Direct Submission  
 Submitted (17-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 13, 2002 this sequence version replaced g1:18767328.

-----Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 -----Project Information  
 Center project name: GSK2  
 Center clone name: CH230-157L12  
 -----Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye 100% of reads  
 Assembly program: Phrap; version 0.990329  
 Consensus quality: 15986 bases at least Q40  
 Consensus quality: 15668 bases at least Q30  
 Consensus quality: 151388 bases at least Q20  
 -----NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
 \* NOTE: This is a "working draft" sequence. It currently  
 \* consists of 54 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 \* 1  
 \* 1281 1380: contig of 1280 bp in length  
 \* 1381 2779: gap of unknown length  
 \* 2780 2879: contig of 1399 bp in length  
 \* 2880 2880: gap of unknown length  
 \* 4438 4438: contig of 1559 bp in length  
 \* 4539 6012: gap of unknown length  
 \* 6013 6112: contig of 1474 bp in length  
 \* 6112: gap of unknown length



6113 7635: contig of 1523 bp in length  
736 7735: gap of unknown length  
7736 9209: contig of 1474 bp in length  
9210 9309: gap of unknown length  
9310 10427: contig of 1118 bp in length  
10428 10527: gap of unknown length  
10528 11603: contig of 1076 bp in length  
11604 11703: gap of unknown length  
11704 13101: contig of 1398 bp in length  
13102 13201: gap of unknown length  
13202 14668: contig of 1467 bp in length  
14669 14768: gap of unknown length  
14769 15842: contig of 1074 bp in length  
15843 15942: gap of unknown length  
15943 17269: contig of 1327 bp in length  
17270 17369: gap of unknown length  
17370 18936: contig of 1567 bp in length  
18937 19036: gap of unknown length  
19037 20110: contig of 1074 bp in length  
20111 20210: gap of unknown length  
20211 22223: contig of 2013 bp in length  
22224 22323: gap of unknown length  
22324 23473: contig of 1150 bp in length  
23474 23573: gap of unknown length  
23574 24899: contig of 1326 bp in length  
24900 24999: gap of unknown length  
25000 26247: contig of 1248 bp in length  
26248 26347: gap of unknown length  
26348 27995: contig of 1648 bp in length  
27996 28095: gap of unknown length  
28096 29735: contig of 1640 bp in length  
29736 29835: gap of unknown length  
29836 31035: contig of 1200 bp in length  
31036 31135: gap of unknown length  
31136 32181: contig of 1046 bp in length  
32182 32281: gap of unknown length  
32282 34586: contig of 2305 bp in length  
34587 34686: gap of unknown length  
34687 35875: contig of 1189 bp in length  
35876 35975: gap of unknown length  
35976 37133: contig of 1158 bp in length  
37134 37233: gap of unknown length  
37234 39149: contig of 1916 bp in length  
39150 39249: gap of unknown length  
39250 41559: contig of 2310 bp in length  
41560 41659: gap of unknown length  
41660 43823: contig of 2164 bp in length  
43824 43923: gap of unknown length  
43924 46913: contig of 2990 bp in length  
46914 47013: gap of unknown length  
47014 49542: contig of 2529 bp in length  
49543 49642: gap of unknown length  
49643 51523: contig of 1881 bp in length  
51524 51623: gap of unknown length  
51623 53986: contig of 2363 bp in length  
53987 54086: gap of unknown length  
54087 57318: contig of 3232 bp in length  
57319 57418: gap of unknown length  
57419 60601: contig of 3183 bp in length  
60602 60701: gap of unknown length  
60702 63800: contig of 3099 bp in length  
63801 63900: gap of unknown length  
63901 67778: contig of 3878 bp in length  
67779 67878: gap of unknown length  
67879 70679: contig of 2801 bp in length  
70680 70779: gap of unknown length  
70780 73224: contig of 3145 bp in length  
73225 74024: gap of unknown length  
74025 77919: contig of 3895 bp in length  
77920 78019: gap of unknown length  
78020 83314: contig of 5295 bp in length  
83315 83414: gap of unknown length  
83415 87408: contig of 3994 bp in length

87409 87508: gap of unknown length  
87509 91364: contig of 3856 bp in length  
91365 91464: gap of unknown length  
91465 95033: contig of 3569 bp in length  
95034 95133: gap of unknown length  
95134 98981: contig of 3848 bp in length  
98982 99082: gap of unknown length  
99082 104913: contig of 5832 bp in length  
104914 105013: gap of unknown length  
105014 110056: contig of 5043 bp in length  
110057 110156: gap of unknown length  
110157 116497: contig of 6341 bp in length  
116498 116597: gap of unknown length  
116598 121911: contig of 5314 bp in length  
121912 122011: gap of unknown length  
122012 127568: contig of 5557 bp in length  
127569 127668: gap of unknown length  
127669 134565: contig of 6897 bp in length  
134566 134665: gap of unknown length  
134666 145980: contig of 11315 bp in length  
145981 146080: gap of unknown length  
146081 154722: contig of 8642 bp in length  
154723 154822: gap of unknown length  
154823 166077: contig of 11255 bp in length  
166078 166177: gap of unknown length  
166178 186415: contig of 20238 bp in length.

Query Match 59.3% Score 492; DB 2; Length 186415;  
Best Local Similarity 80.6%; Pred. No. 1e-128;  
Matches 651; Conservative 0; Mismatches 120; Indels 37; Gaps 5;

QY 6 TTGGCATTCACAGTGGCTCTCTTCACCTTCTCCGACGCCCTGGTCCCAAG 65  
DB 161833 TTGGATTCACCTTGCTGCTTACCTTCTCTGCTCCACGATGATCCCAAG 161774

QY 66 CAGAGAAATACCTGCTGGTGGACCCCTCTCCATGAAACAGAGATGATGATG 125  
DB 161773 AGCGCGCAT---CTGGCAGAGCGCTCGACAGACCAGAGATGATGATG 161717

QY 126 TCAAGACCTATGATGATGATGATGATGATGATGATGATGATGATGATG 185  
DB 161716 TCAAGACCTATGATGATGATGATGATGATGATGATGATGATGATGATG 161657

QY 186 CCACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 245  
DB 161656 CTACATGATGATGATGATGATGATGATGATGATGATGATGATGATG 161597

QY 246 TCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 305  
DB 161596 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 161537

QY 306 CTTCACGAGTGTCCAAATCAGCGTGTGATGATGATGATGATGATGATG 365  
DB 161536 CTTCAGTGTGTCCAAATCAGCGTGTGATGATGATGATGATGATGATG 161477

QY 366 CAGGCAGACGAGCGCTGTGATGATGATGATGATGATGATGATGATGATG 425  
DB 161476 CAGGCAGACGAGCGCTGTGATGATGATGATGATGATGATGATGATGATG 161417

QY 426 AACGGAGCTTCAACAGAACTTCAATCCATCCAGCCAGCCAGCCAGCCAG 485  
DB 161416 AACGGAGCTTCAACAGAACTTCAATCCATCCAGCCAGCCAGCCAGCCAG 161357

QY 486 GGAAGATGCAAGTACAGCTCCGCTTCAGAGATGAGCCAGCTAGTGGGCTG 545  
DB 161356 GGAAGATGCAAGTACAGCTCCGCTTCAGAGATGAGCCAGCTAGTGGGCTG 161297

QY 546 GGTCTCTTGTCTTCAATTCAGCCAGCTCTGATGATGATGATGATGATGAT 605  
DB 161296 GGTCTCTTGTCTTCAATTCAGCCAGCTCTGATGATGATGATGATGATG 161243

QY 606 ATGTGAGAGACAACTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 665  
DB 161243 ATGTGAGAGACAACTTCTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 161183

Db 161242 AGCTGAGAACACTAGTTCATTTTGTGTGCACAGCAGACTGCACAAC----- 161195

QY 666 AGTTTGAACTTTTGTGTGCACAGTTCAGAGAGCCCTTCTCAATATATACTTGT 725

Db 161194 -----TGGAGTTCGACTCCTAGAACCCCTTCTCTGTATATAACGCT 161150

QY 726 TTGTCATAGTTCCTTTCACAGTAACTTTCTATTTCTACTTGCAGTAGAG 785

Db 161149 TTTCATAGTTCCTTTCACAGTAACTTTCTATTTCTACTTGCAGTAGAG 161091

QY 786 ACTCTGATTCCTGGAATTCAGCAATA 813

Db 161090 AACCTGTTCTGGAATTCAGCAATA 161063

RESULT 11

AF156958 644 bp mRNA linear ROD 07-DEC-2000

LOCUS AF156958

DEFINITION Mus musculus NTF2-related export protein NXT1 (NXT1) mRNA, complete cds.

ACCESSION AF156958

VERSION AF156958.2

KEYWORDS GI:11597237

SOURCE Mus musculus.

ORGANISM Mus musculus.

REFERENCE 1 (bases 1 to 644)

AUTHORS Black,B.E., Levesque,L., Holaska,J.M., Wood,T.C. and Paschal,B.M.

TITLE Identification of an NTF2-related factor that binds Ran-GTP and regulates nuclear protein export

JOURNAL Mol. Cell. Biol. 19 (12), 8616-8624 (1999)

MEDLINE 20036817

PUBMED 10567585

REFERENCE 2 (bases 1 to 644)

AUTHORS Black,B.E. and Paschal,B.M.

TITLE Direct Submission

JOURNAL Submitted (09-JUN-1999) Center for Cell Signaling, University of Virginia, 7161 Hospital West, Box 577 Health Sciences Center, Charlottesville, VA 22908, USA

REFERENCE 3 (bases 1 to 644)

AUTHORS Black,B.E. and Paschal,B.M.

TITLE Direct Submission

JOURNAL Submitted (07-DEC-2000) Center for Cell Signaling, University of Virginia, 7161 Hospital West, Box 577 Health Sciences Center, Charlottesville, VA 22908, USA

REMARK COMMENT On Dec 7, 2000 this sequence version replaced gi.5880866.

FEATURES

SOURCE

1. 644

Location/Qualifiers

1. 644

/organism="Mus musculus"

/db\_xref="taxon:10090"

1. 644

/gene="NXT1"

222. 644

/gene="NXT1"

/function="stimulates nuclear protein export"

/note="Similar to the sequence presented in GenBank Accession Number AAA5905; nuclear transport factor; binds Ran-GTP"

/codon\_start=1

/product="NTF2-related export protein NXT1"

/protein\_id="AA054943.2"

/db\_xref="GI:11597238"

/translation="MASVDEKTYVDQACRAAEFVNYVTMDKRRRLSLRYMGATLVNNGAVSDESLSEFEMLPSEFOISYVDCOPVDDATPSQTYLVAVICGVKKEGKRDENONFILTAAQSPNTWTKIASDCEFRQDMAS"

BASE COUNT 150 a 168 c 192 g 134 t

Query Match 45.9%; Score 381.2; DB 10; Length 644;

Best Local Similarity 87.7%; Pred. No. 2,3e-97;

Matches 420; Conservative 0; Mismatches 58; Indels 2; Gaps 1;

QY 45 GCAGCCCTGTTTCCCAAGGACAGAGAAATACCTGCTGAGACCTCTCCATAGAC 104

Db 159 GCAGTCCCTGATCCCTAAGAGAGAGAGCGCT -CAGCGGAGAGCGCTCTTGTCCAGCAAC 216

QY 105 CAGAGATGGCATCTGTGATTTCAAGACCTATGTGATACAGGCTTCAGAGAGCTGAGG 164

Db 217 CAGAGATGGCATCTGTGATTTCAAGACCTATGTGATACAGGCTTCAGAGAGCTGAGG 276

QY 165 AGTTGTCAATGCTCTACTACACGACATGATGATGAGCGGCGCTTTCCTCCGCTGT 224

Db 277 AATTGTCAATGCTCTACTACACGACATGATGATGAGCGGCGCTTTCCTCCGCTGT 336

QY 225 ACATGGGACACAGCCACCTGCTGTGATGATGATGATGATGATGATGATGATGATGAT 284

Db 337 ACATGGGACACAGCCACCTGCTGTGATGATGATGATGATGATGATGATGATGATGAT 396

QY 285 GTGAGTTTGTGAAATGCTGCTTCCAGCGAGTTCCTCAACACAGCGTGTGAGTCCAGC 344

Db 397 GTGAGTTTGTGAGATGCTTCCAGCGAGTTCCTCAACACAGCGTGTGAGTCCAGC 456

QY 345 CTGTCATGATGAGACCCACACAGCCAGACAGCGTCTTGTGATCTGTGATCAG 404

Db 457 CTGTCATGATGAGACCCACACAGCCAGACAGCGTCTTGTGATCTGTGATCAG 516

QY 405 TGAAGTTTGAAGGAGAACCAACAGGACTTCAACGAACTTCATCTGACCGCCAGG 464

Db 517 TGAAGTTTGAAGGAGAACCAACAGGACTTCAACGAACTTCATCTGACCGCCAGG 576

QY 465 CCFACCCACACACAGGAGTGAAGATGCAAGTACTGCTTCCGCTTCAGAGACTGGG 524

Db 577 CGTACCCACAGGAGGCTGTGAGAGATGCAAGTACTGCTTCCGCTTCAGAGACTGGG 636

QY 525 CCAGCTAG 532

Db 637 CCAGCTAG 644

RESULT 12

AC128310/c

LOCUS AC128310

DEFINITION Rattus norvegicus clone CH230-326L8, \*\*\* SEQUENCING IN PROGRESS

ACCESSION AC128310.1

VERSION AC128310.1

KEYWORDS HTG; HTGS; PHASE1.

SOURCE Rattus norvegicus.

ORGANISM Rattus norvegicus.

REFERENCE 1 (bases 1 to 182118)

AUTHORS Muzny,D.M., Adams,C., Adlo-Oduola,B., Ali-Osman,F.R., Allen,C., Alshbrook,S.L., Amaratunga,H.C., Are,J.R., Ayale,M., Banks,T., Barbarella,J., Benton,J., Bimarge,K., Blankenburg,K., Bonnin,D., Bouck,J., Boyle,S., Brieve,M., Brown,E., Brown,M., Bryant,N.P., Buhaig,J., Butch,P., Burkett,C., Burrell,K.L., Byrd,N.C., Carton,T.F., Carter,M., Cavazos,S.R., Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Z., Chowdhury,I., Christopoulos,C., Cleveland,C.D., Cox,C., Coyle,M.D., Dahorne,S.R., David,R., Davila,M.L., Davis,C., Davy-Carroll,L., Dederich,D.A., Delaney,K.R., Delgado,O., Denn,A.L., Ding,Y., Dinh,H.H., Douthwaite,K.J., Draper,H., Dugan-Rocha,S., Durbin,K.J., Earhart,K.C., Edgar,D., Edwards,C.C., Elhaj,C., Escoto,M., Falls,T., Ferraguto,D., Flagg,N., Ford,J., Foster,P., Frantz,P., Gabisi,A., Gao,J., Garcia,A., Garner,T., Garza,N., Gill,R., Gorrell,J.H., Guevara,M., Gunaratne,P., Hale,S., Hamilton,K., Harris,C., Harris,K., Hart,M., Havlak,B., Hawes,C., Hernandez,J., Hernandez,O., Hodgson,A., Hogues,M., Hollway,C., Hollins,B., Homsl,F., Howard,S., Huber,J., Huyl,S.S., Hume,J., Jackson,L.E., Jacobson,B., Jia,Y., Johnson,R., Jolivet,S., Joudah,S., Karlsson,E., Kelly,S., Khan,U., King,L., Korvah,J., Kovar,C., Kratovic,J., Kureshi,A., Landry,N., Leal,B., Lewis,L.C., Lewis,L.,



TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

## COMMENT

Li, J., Li, Z., Lichtarge, O., Liu, C., Liu, J., Liu, W., Lou, S., Lou, S.,  
Lozano, R., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
Maheshwari, M., Marpa, P., Martin, R., Martindale, A., Martinez, E.,  
Massey, E., Mawhney, E., McLeod, M., Meador, K., Meier, M., Metker, N.,  
Miner, G., Miner, Z., Mitchell, T., Mohabbat, K., Morgan, M., Morris, S.,  
Moser, N., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
Nguyen, N., Nickerson, E., Nwokenwo, S., Ogun, M., Okwuonu, G.,  
Oragunye, N., Oviado, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
Peters, L., Pickens, R., Primus, E., Pu, L., Quiles, M., Ren, Y.,  
Rivers, M., Rojas, A., Rojibokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
Scherrer, S., Scott, G., Shen, H., Shoshitari, N., Slason, I.,  
Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,  
Sutton, A., Swatek, A., Taber, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
Tansey, J., Taylor, C., Taylor, T., Tellod, B., Thomas, N., Thomas, S.,  
Usmani, K., Vasquez, L., Vera, V., Villalón, D., Vinson, R., Wang, Q.,  
Wang, S., Ward-Moore, S., Warren, R., Washington, C., Watlington, S.,  
Williams, G., Williamson, A., Wleczyk, R., Wooden, S., Worley, K.,  
Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorilla, S., Nelson, D.,  
Weinstock, G., and Gibbs, R.

Direct Submission  
Unpublished  
2 (bases 1 to 182118)  
Worley, K.C.  
Submitted (19-JUL-2002) Human Genome Sequencing Center, Department  
of Molecular and Human Genetics, Baylor College of Medicine, One  
Baylor Plaza, Houston, TX 77030, USA

----- Genome Center of Medicine -----  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
Project information  
Center project name: GYDN  
Center clone name: CH230-326L8  
Summary Statistics  
Sequencing vector: Plasmid  
Chemistry: Dye-terminator Big Dye 100% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 118171 bases at least Q40  
Consensus quality: 125368 bases at least Q30  
Consensus quality: 130285 bases at least Q20

NOTE: Estimated insert size may differ from sequence length  
(see <http://www.hgsc.bcm.tmc.edu/docs/genbankdraftdata.html>).  
NOTE: This is a 'working draft' sequence. It currently  
consists of 69 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 1004: contig of 1004 bp in length  
\* 1005 1104: gap of unknown length  
\* 1105 2235: contig of 1131 bp in length  
\* 2236 2336: gap of unknown length  
\* 2336 3480: contig of 1145 bp in length  
\* 3481 3580: gap of unknown length  
\* 3581 4634: contig of 1054 bp in length  
\* 4634 4734: gap of unknown length  
\* 4735 5901: contig of 1167 bp in length  
\* 5902 6001: gap of unknown length  
\* 6002 7042: contig of 1041 bp in length  
\* 7043 7143: gap of unknown length  
\* 7143 8354: contig of 1212 bp in length  
\* 8355 8455: gap of unknown length  
\* 8455 9464: contig of 1010 bp in length  
\* 9465 9565: gap of unknown length  
\* 9565 10594: contig of 1030 bp in length  
\* 10595 12121: gap of unknown length  
\* 12121 12221: gap of unknown length

12222 13564: contig of 1343 bp in length  
\* 13565 13666: gap of unknown length  
\* 13666 14776: contig of 1112 bp in length  
\* 14777 14876: gap of unknown length  
\* 14877 16124: contig of 1248 bp in length  
\* 16125 16224: gap of unknown length  
\* 16225 17356: contig of 1132 bp in length  
\* 17357 17456: gap of unknown length  
\* 17457 18699: contig of 1243 bp in length  
\* 18700 18799: gap of unknown length  
\* 18800 20696: contig of 1897 bp in length  
\* 20697 20797: gap of unknown length  
\* 20797 21843: contig of 1047 bp in length  
\* 21844 21943: gap of unknown length  
\* 21944 23254: contig of 1311 bp in length  
\* 23255 23354: gap of unknown length  
\* 23355 24597: contig of 1243 bp in length  
\* 24598 24697: gap of unknown length  
\* 24698 26076: contig of 1379 bp in length  
\* 26077 26176: gap of unknown length  
\* 26177 27645: contig of 1469 bp in length  
\* 27646 27745: gap of unknown length  
\* 27746 29460: contig of 1715 bp in length  
\* 29461 29560: gap of unknown length  
\* 29561 30694: contig of 1134 bp in length  
\* 30695 30794: gap of unknown length  
\* 30795 32923: contig of 2129 bp in length  
\* 32924 33023: gap of unknown length  
\* 33024 34861: contig of 1838 bp in length  
\* 34862 34961: gap of unknown length  
\* 34962 37952: contig of 2991 bp in length  
\* 37953 38052: gap of unknown length  
\* 38053 40327: contig of 2175 bp in length  
\* 40328 40327: gap of unknown length  
\* 40328 42768: contig of 2441 bp in length  
\* 42769 42868: gap of unknown length  
\* 42869 44526: contig of 1658 bp in length  
\* 44527 44626: gap of unknown length  
\* 44627 46017: contig of 1391 bp in length  
\* 46018 46117: gap of unknown length  
\* 46118 48882: contig of 2765 bp in length  
\* 48883 48982: gap of unknown length  
\* 48983 50447: contig of 1465 bp in length  
\* 50448 50547: gap of unknown length  
\* 50548 52563: contig of 2016 bp in length  
\* 52564 52663: gap of unknown length  
\* 52664 53812: contig of 1149 bp in length  
\* 53813 53912: gap of unknown length  
\* 53913 55339: contig of 1337 bp in length  
\* 55340 57051: gap of unknown length  
\* 57052 57151: contig of 1712 bp in length  
\* 57152 58434: contig of 1283 bp in length  
\* 58435 58534: gap of unknown length  
\* 58535 59907: contig of 1373 bp in length  
\* 59908 60008: gap of unknown length  
\* 60009 62417: contig of 2410 bp in length  
\* 62418 62517: gap of unknown length  
\* 62519 65195: contig of 2678 bp in length  
\* 65196 65295: gap of unknown length  
\* 65296 67509: contig of 2214 bp in length  
\* 67510 67609: gap of unknown length  
\* 67610 69530: contig of 1921 bp in length  
\* 69531 71223: contig of 2093 bp in length  
\* 71224 71823: gap of unknown length  
\* 71824 74364: contig of 2441 bp in length  
\* 74365 74364: gap of unknown length  
\* 74365 76765: contig of 2401 bp in length  
\* 76766 76865: gap of unknown length  
\* 76866 79018: contig of 2153 bp in length  
\* 79019 82569: contig of 3451 bp in length

82570 82669: gap of unknown length  
 \* 82670 84299: contig of 1630 bp in length  
 \* 84300 84399: gap of unknown length  
 \* 84400 87075: contig of 2676 bp in length  
 \* 87076 87175: gap of unknown length  
 \* 87176 91309: contig of 4134 bp in length  
 \* 91310 91409: gap of unknown length  
 \* 91410 93481: contig of 2072 bp in length  
 \* 93482 93582: gap of unknown length  
 \* 93583 97220: contig of 3638 bp in length  
 \* 97221 97319: gap of unknown length  
 \* 97320 100038: contig of 2719 bp in length  
 \* 100039 100138: gap of unknown length  
 \* 100139 103847: contig of 3709 bp in length  
 \* 103848 103948: gap of unknown length  
 \* 103949 107670: contig of 3723 bp in length  
 \* 107671 107770: gap of unknown length  
 \* 107771 110595: contig of 2825 bp in length  
 \* 110596 110696: gap of unknown length  
 \* 110697 115142: contig of 4447 bp in length  
 \* 115143 115242: gap of unknown length

Query Match 43.1%; Score 357.8; DB 2; Length 182118;  
 Best Local Similarity 81.2%; Pred. No. 1.8e-90;  
 Matches 478; Conservative 0; Mismatches 77; Indels 34; Gaps 4;

225 ACATGGGACACGACCCCTGCTGGAATGCAATGCTGTTTCAGAGACAAATCTTGA 284  
 Db ACATGGGACACGACCCCTGCTGGAATGCAATGCTGTTTCAGAGACAAATCTTGA 32838  
 285 GTGAGTTTGTGAATGTTGCTTCCAGCGAGTTCACAAATCAGCGGTGAGCTCCAGC 344  
 Db GTGAGTTTGTGAATGTTGCTTCCAGCGAGTTCACAAATCAGCGGTGAGCTCCAGC 328378  
 32837 GTGAGTTTGTGAATGTTGCTTCCAGCGAGTTCACAAATCAGCGGTGAGCTCCAGC 327778  
 345 CTGTTTCATGATGAGACCCACACAGCCAGCAGCGGTCTGTTGTCATCTGTGATGAG 404  
 Db CTGTTTCATGATGAGACCCACACAGCCAGCAGCGGTCTGTTGTCATCTGTGATGAG 327777  
 32777 CTGTTTCATGATGAGACCCACACAGCCAGCAGCGGTCTGTTGTCATCTGTGATGAG 32718  
 405 TGAAGTTTGAAGGAGAACAAACAGCGGACTTCACAGAACTTCTGACCGCCAGC 464  
 Db TGAAGTTTGAAGGAGAACAAACAGCGGACTTCACAGAACTTCTGACCGCCAGC 32717  
 32717 TGAAGTTTGAAGGAGAACAAACAGCGGACTTCACAGAACTTCTGACCGCCAGC 32658  
 465 CCTACACCCAGAACACACTGTGGAAGATGCAAGTACTGCTCCGCTCCAGGACTGG 524  
 Db CCTACACCCAGAACACACTGTGGAAGATGCAAGTACTGCTCCGCTCCAGGACTGG 32657  
 32657 CGTACACCCAGAACACACTGTGGAAGATGCAAGTACTGCTCCGCTCCAGGACTGG 32598  
 525 CCAGCTAGTGGGGGTGGCAGAGGCTCTTGTCTATTCAGCCCTAGCTGTAGAGAAA 584  
 Db CCAGCTAGTGGGGGTGGCAGAGGCTCTTGTCTATTCAGCCCTAGCTGTAGAGAAA 32597  
 32597 CCAGCTAGTGGGGGTGGCAGAGGCTCTTGTCTATTCAGCCCTAGCTGTAGAGAAA 32543  
 585 TGCACACCTGCACTCTCAAGATGTGAGGAACACAGTTCATTTCTGTTGTCGAGAGC 644  
 Db TGCACACCTGCACTCTCAAGATGTGAGGAACACAGTTCATTTCTGTTGTCGAGAGC 32542  
 32542 TGCCAGCTCAAAATCTC-AGGACGTGAGAGAACACTGATTTTGTGTGACAGAGC 32484  
 645 ACTGACAGCTCAGCTGTCGCCAGGTGAACCTTTTGTGTGCTCAAGTCTAGAGAGTCC 704  
 Db ACTGACAGCTCAGCTGTCGCCAGGTGAACCTTTTGTGTGCTCAAGTCTAGAGAGTCC 32483  
 32483 ACTGACAGCTCAGCTGTCGCCAGGTGAACCTTTTGTGTGCTCAAGTCTAGAGAGTCC 32451  
 705 CTTCCTCAATATATATCTGTTGTCATAGTTCCTTTTCAAGTAGTAACCTTTCTAT 764  
 Db CTTCCTCAATATATATCTGTTGTCATAGTTCCTTTTCAAGTAGTAACCTTTCTAT 32450  
 32450 CTTCCTCAATATATATCTGTTGTCATAGTTCCTTTTCAAGTAGTAACCTTTCTAT 32392  
 765 TTTTCTACTTGCCAGTAGAGACTGTGATCTGGAATTTCTGACAAATA 813  
 Db TTTTCTACTTGCCAGTAGAGACTGTGATCTGGAATTTCTGACAAATA 32391  
 32391 TTTTCTACTTGCCAGTAGAGACTGTGATCTGGAATTTCTGACAAATA 32343

RESULT 13  
 AX261550 318 bp DNA linear PAT 26-OCT-2001  
 LOCUS AX261550  
 DEFINITION Sequence 1201 from Patent WO01/73027.  
 ACCESSION AX261550  
 VERSION AX261550.1 GI:15510517

KEYWORDS  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Carnivora; Homnidae; Homo.  
 REFERENCE  
 1 Meagher, M.J., Xu, J. and King, G.E.  
 TITLES  
 JOURNAL Composition and methods for therapy and diagnosis of colon cancer  
 PATENT: WO 01/73027-A 1201 04-OCT-2001;  
 CORNIX CORPORATION (US)  
 FEATURES  
 Source Location/Qualifiers  
 1. 318  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 BASE COUNT 76 a 82 c 83 g 76 t 1 others  
 ORIGIN

Query Match 38.2%; Score 317; DB 6; Length 318;  
 Best Local Similarity 99.7%; Pred. No. 4.3e-79;  
 Matches 317; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

147 CCTGACAGCTGCTGAGAGATTTGTCATGCTCTACTACACCATGATAGCGCGGC 206  
 Db CCTGACAGCTGCTGAGAGATTTGTCATGCTCTACTACACCATGATAGCGCGGC 60  
 1 CCTGACAGCTGCTGAGAGATTTGTCATGCTCTACTACACCATGATAGCGCGGC 207  
 207 GTTGGCTGCTCCGCTGTACATGAGGACACACCCCTGCTGGAATGCAATGCTGTT 266  
 Db GTTGGCTGCTCCGCTGTACATGAGGACACACCCCTGCTGGAATGCAATGCTGTT 120  
 61 GTTGGCTGCTCCGCTGTACATGAGGACACACCCCTGCTGGAATGCAATGCTGTT 120  
 267 CAGACAAAGAAATCCTTGAAGTAGTTTGTGAATGTTCCCTCCAGCAGTTCAAATCA 326  
 Db CAGACAAAGAAATCCTTGAAGTAGTTTGTGAATGTTCCCTCCAGCAGTTCAAATCA 121  
 121 CAGACAAAGAAATCCTTGAAGTAGTTTGTGAATGTTCCCTCCAGCAGTTCAAATCA 180  
 327 GCGTGTGAGACTGCCAGCCTGTTCAATGATGAAGACCAACAGCCAGCAGCGTCTG 386  
 Db GCGTGTGAGACTGCCAGCCTGTTCAATGATGAAGACCAACAGCCAGCAGCGTCTG 181  
 181 GCGTGTGAGACTGCCAGCCTGTTCAATGATGAAGACCAACAGCCAGCAGCGTCTG 240  
 387 TTGTATCTGTGATGATGAGAGCTTGAAGGAGAACAAACAGGACTTCAACAGAACT 446  
 Db TTGTATCTGTGATGATGAGAGCTTGAAGGAGAACAAACAGGACTTCAACAGAACT 241  
 241 TTGTATCTGTGATGATGAGAGCTTGAAGGAGAACAAACAGGACTTCAACAGAACT 300  
 447 TCATCCTGACCGCCAGC 464  
 Db TCATCCTGACCGCCAGC 301  
 301 TCATCCTGACCGCCAGC 318

RESULT 14  
 AC11510/c  
 LOCUS AC11510  
 DEFINITION Rattus norvegicus clone CH230-133M15, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\* 58 unordered pieces:  
 AC11510  
 AC11510.2 GI:21736059  
 HTG: HTGS PHASE1.  
 KEYWORDS  
 SOURCE Norway rat.  
 ORGANISM Rattus norvegicus  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;  
 Rattus.  
 1 (bases 1 to 153483)  
 REFERENCE  
 Muzny, D.M., Adams, C., Adio-Oduola, B., All-ouman, F.R., Allen, C.,  
 Albrooks, S.L., Amaratunga, H.C., Are, J.R., Ayele, M., Banks, T.,  
 Barbarella, J., Benton, J., Blinze, K., Blankenburg, K., Bonin, D.,  
 Bouck, J., Bowler, S., Brileva, M., Brown, E., Brown, M., Bryant, N.P.,  
 Buhay, C., Burch, P., Burkett, C., Burrell, K.L., Byrd, N.C.,  
 Caron, T.F., Carter, M., Cavazos, S.R., Chacko, J., Chavez, D.,  
 Chen, G., Chen, Z., Chowdhury, I., Christopoulos, C.,  
 Cleveland, C.D., Cox, C., Coyle, M.D., Dathorne, S.R., David, R.,  
 Davila, M.L., Davis, C., Davy-Carroll, L., Dederich, D.A.,  
 Delaney, K.R., Delgado, O., Denn, A.L., Ding, Y., Dinh, H.H.,  
 Dunham, K.J., Draper, H., Dugan-Rocha, S., Durbin, K.J.,  
 Earhart, C., Edgar, D., Edwards, C.C., Elhaj, C., Escotto, M.,  
 Falls, T., Ferraguto, D., Flagg, N., Ford, J., Foster, P., Frantz, P.,

Gabist, A., Gao, J., Garcia, A., Garner, T., Garza, N., Gill, R.,  
 Gori, J. H., Guevara, W., Gunaratne, P., Hale, S., Hamilton, K.,  
 Harris, C., Harris, R., Hart, M., Havlik, P., Hawes, A., Hernandez, J.,  
 Hernandez, O., Hodgson, A., Hogue, M., Hollaway, C., Hollins, B.,  
 Homs, F., Howard, S., Huber, J., Hult, S., Hume, J., Jackson, L. E.,  
 Jacobson, B., Jia, Y., Johnson, R., Jolivet, S., Joudah, S.,  
 Karlsson, E., Kelly, S., Khan, U., King, L., Korvah, J., Kovar, C.,  
 Kratoch, J., Kureshi, A., Landry, N., Leal, B., Lewis, L. C., Lewis, L.,  
 Li, J., Li, Z., Licharge, O., Lien, C., Liu, J., Liu, W., Louised, H.,  
 Lozano, R. J., Lu, X., Lucier, A., Lucier, R., Luna, R., Ma, J.,  
 Maheshwari, M., Mapa, P., Martin, R., Martindale, A., Martinez, E.,  
 Massey, E., Mawhinney, E., McLeod, M. P., Meador, M., Mei, G., Metzger, M.,  
 Moser, M., Neal, D., Newton, J., Newton, N., Nguyen, A., Nguyen, N.,  
 Nguyen, N., Nickerson, E., Nwokweto, S., Ogun, M., Okunolu, G.,  
 Otaguaye, N., Oviedo, R., Pace, A., Payton, B., Peery, J., Perez, L.,  
 Peters, L., Pickens, R., Primus, E., Pu, L. L., Quiles, M., Ren, Y.,  
 Rivers, M., Rojas, A., Rojebokan, I., Rolfe, M., Ruiz, S., Savary, G.,  
 Scherer, S., Scott, G., Shen, H., Shooshari, N., Sisson, I.,  
 Sodergren, E., Sonake, T., Sparks, A., Stanley, H., Stone, H.,  
 Sutton, A., Swatek, A., Tabor, P., Tamerisa, A., Tamerisa, K., Tang, H.,  
 Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N., Thomas, S.,  
 Usmani, K., Vasquez, L., Vera, Y., Villalón, D., Vinson, R., Wang, O.,  
 Wang, S., Ward-Moore, S., Warren, R., Washington, C., Wellington, S.,  
 Williams, G., Williamson, A., Wleczek, R., Woodden, S., Worley, K.,  
 Wu, C., Wu, Y., Wu, Y. F., Zhou, J., Zorrilla, S., Nelson, D.,  
 Weinstein, G., and Gibbs, R.  
 Unpublished  
 Direct Submission  
 2 (bases 1 to 153483)  
 Worley, K. C.  
 Submitted (19-FEB-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 3 (bases 1 to 153483)  
 Worley, K. C.  
 Direct Submission  
 Submitted (13-JUL-2002) Human Genome Sequencing Center, Department  
 of Molecular and Human Genetics, Baylor College of Medicine, One  
 Baylor Plaza, Houston, TX 77030, USA  
 On Jul 12, 2002 this sequence version replaced g1:18701275.  
 ----- Genome Center  
 Center: Baylor College of Medicine  
 Center code: BCM  
 Web site: <http://www.hgsc.bcm.tmc.edu/>  
 Contact: hgsc-help@bcm.tmc.edu  
 ----- Project Information  
 Project name: GMLT  
 Center project name: CH230-133N15  
 ----- Summary Statistics  
 Sequencing vector: Plasmid  
 Chemistry: Dye-terminator Big Dye 100% of reads  
 Assembly program: Phrap: version 0.990329  
 Consensus quality: 88643 bases at least Q40  
 Consensus quality: 93187 bases at least Q30  
 Consensus quality: 97900 bases at least Q20  
 -----  
 \* NOTE: Estimated insert size may differ from sequence length  
 \* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_data.html)).  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 58 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.  
 1 1313: contig of 1313 bp in length  
 \* 1314 1413: gap of unknown length  
 \* 3169 3168: contig of 1755 bp in length  
 \* 3269 4610: gap of unknown length  
 4610: contig of 1342 bp in length  
 4611 4710: gap of unknown length  
 \* 4711 6083: contig of 1373 bp in length  
 \* 6084 6184: gap of unknown length  
 \* 6185 7726: contig of 1542 bp in length  
 \* 7727 7826: gap of unknown length  
 \* 7827 8962: contig of 1137 bp in length  
 \* 8963 9062: gap of unknown length  
 \* 9063 10153: contig of 1090 bp in length  
 \* 10154 10253: gap of unknown length  
 \* 10254 11464: contig of 1211 bp in length  
 \* 11465 11564: gap of unknown length  
 \* 11565 13182: contig of 1619 bp in length  
 \* 13183 13283: gap of unknown length  
 \* 13284 14545: contig of 1263 bp in length  
 \* 14546 15847: gap of unknown length  
 \* 15848 15947: contig of 1202 bp in length  
 \* 15948 17592: gap of unknown length  
 \* 17593 17692: contig of 1645 bp in length  
 \* 17693 18890: gap of unknown length  
 \* 18891 18990: contig of 1198 bp in length  
 \* 18991 20761: gap of unknown length  
 \* 20762 20860: contig of 1770 bp in length  
 \* 20861 22271: gap of unknown length  
 \* 22272 22372: contig of 1411 bp in length  
 \* 22373 23809: gap of unknown length  
 \* 23810 25341: contig of 1438 bp in length  
 \* 25342 25441: gap of unknown length  
 \* 25442 26863: contig of 1422 bp in length  
 \* 26864 26963: gap of unknown length  
 \* 26964 28125: contig of 1162 bp in length  
 \* 28126 28225: gap of unknown length  
 \* 28226 29789: contig of 1564 bp in length  
 \* 29790 29880: gap of unknown length  
 \* 29881 31582: contig of 1693 bp in length  
 \* 31583 32773: gap of unknown length  
 \* 32774 32873: contig of 1091 bp in length  
 \* 32874 34244: gap of unknown length  
 \* 34245 34344: contig of 1371 bp in length  
 \* 34345 36071: gap of unknown length  
 \* 36072 36171: contig of 1727 bp in length  
 \* 36172 37383: gap of unknown length  
 \* 37384 37483: contig of 1212 bp in length  
 \* 37484 38840: gap of unknown length  
 \* 38841 38940: contig of 1357 bp in length  
 \* 38941 40471: gap of unknown length  
 \* 40472 40571: contig of 1531 bp in length  
 \* 40572 42670: gap of unknown length  
 \* 42671 42770: contig of 2099 bp in length  
 \* 42771 45260: gap of unknown length  
 \* 45261 45360: contig of 2490 bp in length  
 \* 45361 47344: gap of unknown length  
 \* 47345 49411: contig of 1984 bp in length  
 \* 49412 49511: gap of unknown length  
 \* 49512 51405: contig of 1967 bp in length  
 \* 51406 51505: gap of unknown length  
 \* 51506 53397: contig of 1894 bp in length  
 \* 53398 53497: gap of unknown length  
 \* 53498 56170: contig of 2673 bp in length  
 \* 56171 56270: gap of unknown length  
 \* 56271 59034: contig of 2764 bp in length  
 \* 59035 59134: gap of unknown length  
 \* 59135 60990: contig of 1856 bp in length  
 \* 60991 61090: gap of unknown length  
 \* 61091 62910: contig of 1820 bp in length  
 \* 62911 63010: gap of unknown length  
 \* 63011 65846: contig of 2336 bp in length  
 \* 65847 67736: gap of unknown length  
 \* 67737 67836: contig of 2090 bp in length  
 67836: gap of unknown length

COMMENT

REFERENCE  
AUTHORS  
TITLE  
JOURNAL

JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL

```

* 67837 70060: contig of 2224 bp in length
* 70061 70160: gap of unknown length
* 70161 72709: contig of 2549 bp in length
* 72710 72809: gap of unknown length
* 72810 75211: contig of 2402 bp in length
* 75211 75311: gap of unknown length
* 75312 77837: contig of 2526 bp in length
* 77838 77937: gap of unknown length
* 77938 80801: contig of 2664 bp in length
* 80802 80902: gap of unknown length
* 80903 85672: contig of 4771 bp in length
* 85673 85773: gap of unknown length
* 85774 89247: contig of 3475 bp in length
* 89248 89348: gap of unknown length
* 89349 92449: contig of 3102 bp in length
* 92450 92550: gap of unknown length
* 92551 94576: contig of 2027 bp in length
* 94577 94676: gap of unknown length
* 94677 99771: contig of 4995 bp in length
* 99772 99772: gap of unknown length
* 99773 106874: contig of 7103 bp in length
* 106875 106975: gap of unknown length
* 106976 112478: contig of 5504 bp in length
* 112479 112579: gap of unknown length
* 112580 119343: contig of 6765 bp in length
* 119344 119443: gap of unknown length
* 119444 123855: contig of 4412 bp in length
* 123856 123956: gap of unknown length
* 123957 129007: contig of 5052 bp in length

```

Query Match  
Best Local Similarity 34.7%; Score 287.6; DB 2; Length 153483;  
Matches 381; Conservative 0; Mismatches 179; Indels 6; Gaps 4;

```

6 TTGGCATTACAGTGGCTCTCTTCAACCTTACCTCCGACAGCCCTGGTCCCAAG 65
43232 TGTGATTTACCTTGTGTTCACTTACCTTGTGCTGCTCAGACGATCCCAAG 43173
66 CAGAGAAATACCTGCTGAGGCTCTTCCATGATGAGAGATGATGATGATGAT 125
43172 AGGCGGCAAT---CTGGCAGAGAGGCTCTGCGACAGACAGAGATGATGAT 43116
126 TCAAAACCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 185
43115 TCAAAACCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 43056
186 CCACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 245
43055 CTACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 42996
246 TCTGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 305
42995 TATGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 42936
306 CTTCACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 363
42935 CTTCACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 42876
364 ACCAAGCCAGAGC-ACGATCTGATGATGATGATGATGATGATGATGATGAT 422
42875 GCCAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 42816
423 AACAAAGGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 482
42815 AACAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 42756
483 TGTGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 542
42755 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 42696
543 AGAGGTCTCTTGTCTTCAATGAGCC 568
42695 NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN 42670

```

```

RESULT 15
AF201942
LOCUS Homo sapiens DC9 (DC9) mRNA, complete cds.
DEFINITION AF201942
ACCESSION AF201942
VERSION AF201942.1 GI:9295187
KEYWORDS F11 CDNA.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 896)
AUTHORS Li, Y., Peng, Y., Li, N., Gu, W., Han, Z., Fu, G. and Chen, Z.
TITLE Novel genes expressed in human dendritic cell
JOURNAL Unpublished
2 (bases 1 to 896)
AUTHORS Li, Y., Peng, Y., Li, N., Gu, W., Han, Z., Fu, G. and Chen, Z.
TITLE Submitted (05-NOV-1999) Chinese National Human Genome Center at Shanghai, 351, Guo Shoujing Road, Zhangjiang Hi-Tech Park, Pudong, Shanghai 201203, P. R. China
FEATURES
source 1..896
/organism="Homo sapiens"
/db_xref="taxon:9606"
/cell_type="dendritic cell"
1..896
/gene="DC9"
145..573
/gene="DC9"
/codon_start=1
/product="DC9"
/protein_id="AAF86878.1"
/db_xref="GI:9295188"
/translation="MATSIDPFTYVDACRAAEFVNIYETMDKRRALFRLYLDKA
TIMGNNAVSGDALNFRDPLPSSEFOVNMIDCPVHQAATOSOTTVLVVNSGPVKE
DGNKHFNFONLITRQSTFNNYTKIADSCRRFDMS"
BASE COUNT 267 a 177 c 178 g 274 t
Query Match
Best Local Similarity 30.3%; Score 251.6; DB 9; Length 896;
Matches 326; Conservative 0; Mismatches 124; Indels 0; Gaps 0;
114 CATCTGTGATTTCAAGACCTATGATGATGATGATGATGATGATGATGATGAT 173
152 CGTCTGTGATTTAAACCTATGATGATGATGATGATGATGATGATGATGATG 211
174 ATGCTACTACACACCACCATGATGATGATGATGATGATGATGATGATGATG 223
212 ATATTACTACTGATGATGATGATGATGATGATGATGATGATGATGATGAT 271
234 CAGCCACCCGGCTGGAATGATGATGATGATGATGATGATGATGATGATGAT 293
272 AGGCCACCTTAATTAATGATGATGATGATGATGATGATGATGATGATGAT 331
294 TTGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 353
332 TTACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 391
354 ATGAACCCACACCAAGCAGACAGAGCTCTTGTGATGATGATGATGATGATG 413
392 AGCAACCACTCAGTCCCAACCTACAGTCTTGTGATGATGATGATGATGATG 451
414 AGGGAAACAAACAGGAGATGATGATGATGATGATGATGATGATGATGATG 473
452 ATGGAACAAACAAACATTTCTTCAACAGAGATGATGATGATGATGATGATG 511
474 GCAACACAGTGTGAGAGATGATGATGATGATGATGATGATGATGATGATG 533
512 ACAATCTGTGTGAGAGATGATGATGATGATGATGATGATGATGATGATGAT 571

```

Mon Jun 30 08:51:04 2003

us-09-763-902b-10.rge

Page 18

Oy 534 GGGGTGCGAGGTCCTTTCCTTCATTC 563  
||| ||| ||| ||| ||| ||| |||  
Db 572 AAAGGGCAAAAGTCATTCATTCATTCGTC 601

Search completed: June 28, 2003, 02:38:50  
Job time : 2249.7 secs

---

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: June 24, 2003, 19:20:48 ; Search time 39.6938 Seconds

(without alignments)  
5572.555 Million cell updates/sec

Title: US-09-763-902B-10

Perfect score: 1489

Sequence: 1 tgaacttgcatcactg.....ataattaataacacatg 830

Scoring table:

BLOSUM62	Xgapop 10.0	Xgapext 0.5
	Xgapop 10.0	Xgapext 0.5
	Fgapop 6.0	Fgapext 7.0
	Delop 6.0	Delext 7.0

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 1816940

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Command line parameters:

```

-MODEL=frame_n2p.model -DEV=xlp
-O/cgnet1/USPRO.spool/US09763902/runat_24062003_130340_12405/app.query.fasta.1.2254
-DB-A_Geneseq_101002 -QFWT=fastan -SUFFIX=n2p.rag -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFWT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US09763902.ECGN.1.1.162.gunat_24062003_130340_12405 -NCPU=6 -ICPU=3
-NO.MMAP -LARGESOURCY -NEG.SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6
-Fgapext=7 -Ygapop=10 -Ygapext=0.5 -Delop=6 -Delext=7

```

Database: A\_Geneseq\_101002:\*

```

1: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SID22/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SID22/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	813	54.6	164	AAU16228	Human novel secret
2	732	49.2	140	AAE83318	Human protein tran
3	584.5	39.3	197	AAE95379	Human protein seg
4	584	39.2	142	AAE83323	Human protein tran
5	584	39.2	142	AAE83323	Human NTF2 associa
6	579	38.9	162	AAU16227	Human novel secret
7	308	20.7	133	ABE65017	Drosophila melanog
8	130	8.7	125	AAU16203	Arabidopsis thalia
9	120	8.1	153	AAU16203	Arabidopsis thalia
10	119	8.0	688	ABE67404	Drosophila melanog
11	119	8.0	690	ABE66052	Drosophila melanog
12	117.5	7.9	123	AAU16227	Arabidopsis thalia
13	117.5	7.9	126	AAU16227	Arabidopsis thalia
14	117.5	7.9	131	AAU16227	Arabidopsis thalia
15	117.5	7.9	134	AAU16227	Arabidopsis thalia
16	115.5	7.8	143	AAU16227	Arabidopsis thalia
17	105.5	7.1	130	ABE71855	Arabidopsis thalia
18	105.5	7.1	466	AAU16227	Arabidopsis thalia
19	105.5	7.1	477	AAU16227	Arabidopsis thalia
20	105	7.1	122	AAU16227	Arabidopsis thalia
21	105	7.1	157	AAU16227	Arabidopsis thalia
22	93	6.2	786	AAU16227	Arabidopsis thalia
23	91.5	6.1	508	AAU16227	Arabidopsis thalia
24	91.5	6.1	962	AAU16227	Arabidopsis thalia
25	90.5	6.1	685	AAU16227	Arabidopsis thalia
26	90	6.0	466	AAU16227	Arabidopsis thalia
27	90	6.0	466	AAU16227	Arabidopsis thalia
28	90	6.0	1061	AAU16227	Arabidopsis thalia
29	89.5	6.0	659	AAU16227	Arabidopsis thalia
30	89	6.0	326	AAU16227	Arabidopsis thalia
31	89	6.0	1609	AAU16227	Arabidopsis thalia
32	88.5	5.9	129	AAU16227	Arabidopsis thalia
33	88.5	5.9	130	AAU16227	Arabidopsis thalia
34	88.5	5.9	218	AAU16227	Arabidopsis thalia
35	88.5	5.9	671	AAU16227	Arabidopsis thalia
36	88.5	5.9	671	AAU16227	Arabidopsis thalia
37	88	5.9	1081	AAU16227	Arabidopsis thalia
38	88	5.9	1081	AAU16227	Arabidopsis thalia
39	87.5	5.9	160	AAU16227	Arabidopsis thalia
40	87.5	5.9	160	AAU16227	Arabidopsis thalia
41	87.5	5.9	460	AAU16227	Arabidopsis thalia
42	87.5	5.9	472	AAU16227	Arabidopsis thalia
43	87	5.8	319	AAU16227	Arabidopsis thalia
44	87	5.8	364	AAU16227	Arabidopsis thalia
45	87	5.8	440	AAU16227	Arabidopsis thalia

## ALIGNMENTS

RESULT 1  
AAU16228  
AAU16228 standard; Protein; 164 AA.

AAU16228; 07-NOV-2001 (first entry)

Human novel secreted protein, Seq ID 1181.

Human; immunosuppressive; antiarthritic; antirheumatic;  
cytotoxic; cardiant; vasotropic; cerebroprotective; nootropic;  
neuroprotective; antibacterial; virucide; fungicide; optalmological;  
vulnerary; secreted protein; rheumatoid arthritis;  
hyperproliferative disorder; cardiovascular disorder; cardiac arrest;  
cerebrovascular disorder; cerebral ischemia; angiodenesis;  
nervous system disorder; Alzheimer's disease; infection; ocular disorder;  
corneal infection; wound healing; epithelial cell proliferation;  
skin ageing; food additive; preservative; antiproliferative.

XX Homo sapiens.  
 OS WO20015322-A2.  
 PN 02-AUG-2001.  
 XX 17-JAN-2001: 2001MO-US01341.  
 XX 31-JAN-2000: 2000US-0179065.  
 PR 04-FEB-2000: 2000US-0180628.  
 PR 24-FEB-2000: 2000US-0184664.  
 PR 02-MAR-2000: 2000US-0186350.  
 PR 16-MAR-2000: 2000US-0189874.  
 PR 17-MAR-2000: 2000US-0190076.  
 PR 18-APR-2000: 2000US-0198123.  
 PR 19-MAY-2000: 2000US-0205515.  
 PR 07-JUN-2000: 2000US-0209467.  
 PR 28-JUN-2000: 2000US-0214886.  
 PR 30-JUN-2000: 2000US-0215135.  
 PR 07-JUL-2000: 2000US-0216647.  
 PR 07-JUL-2000: 2000US-0216880.  
 PR 11-JUL-2000: 2000US-0217487.  
 PR 11-JUL-2000: 2000US-0217496.  
 PR 14-JUL-2000: 2000US-0218290.  
 PR 26-JUL-2000: 2000US-0220963.  
 PR 26-JUL-2000: 2000US-0220964.  
 PR 14-AUG-2000: 2000US-0224518.  
 PR 14-AUG-2000: 2000US-0224519.  
 PR 14-AUG-2000: 2000US-0225213.  
 PR 14-AUG-2000: 2000US-0225214.  
 PR 14-AUG-2000: 2000US-0225266.  
 PR 14-AUG-2000: 2000US-0225267.  
 PR 14-AUG-2000: 2000US-0225268.  
 PR 14-AUG-2000: 2000US-0225270.  
 PR 14-AUG-2000: 2000US-0225447.  
 PR 14-AUG-2000: 2000US-0225757.  
 PR 14-AUG-2000: 2000US-0225758.  
 PR 14-AUG-2000: 2000US-0225759.  
 PR 18-AUG-2000: 2000US-0226279.  
 PR 22-AUG-2000: 2000US-0226681.  
 PR 22-AUG-2000: 2000US-0226688.  
 PR 22-AUG-2000: 2000US-0227182.  
 PR 23-AUG-2000: 2000US-0227009.  
 PR 30-AUG-2000: 2000US-0228924.  
 PR 01-SEP-2000: 2000US-0229287.  
 PR 01-SEP-2000: 2000US-0229343.  
 PR 01-SEP-2000: 2000US-0229344.  
 PR 01-SEP-2000: 2000US-0229345.  
 PR 05-SEP-2000: 2000US-0229509.  
 PR 05-SEP-2000: 2000US-0229513.  
 PR 06-SEP-2000: 2000US-0230437.  
 PR 06-SEP-2000: 2000US-0230438.  
 PR 08-SEP-2000: 2000US-0231242.  
 PR 08-SEP-2000: 2000US-0231243.  
 PR 08-SEP-2000: 2000US-0231244.  
 PR 08-SEP-2000: 2000US-0231413.  
 PR 08-SEP-2000: 2000US-0231414.  
 PR 08-SEP-2000: 2000US-0232080.  
 PR 08-SEP-2000: 2000US-0232081.  
 PR 12-SEP-2000: 2000US-0231968.  
 PR 14-SEP-2000: 2000US-0232397.  
 PR 14-SEP-2000: 2000US-0232398.  
 PR 14-SEP-2000: 2000US-0232399.  
 PR 14-SEP-2000: 2000US-0232400.  
 PR 14-SEP-2000: 2000US-0232401.  
 PR 14-SEP-2000: 2000US-0233063.  
 PR 14-SEP-2000: 2000US-0233064.  
 PR 14-SEP-2000: 2000US-0233065.  
 PR 21-SEP-2000: 2000US-0234423.  
 PR 21-SEP-2000: 2000US-0234474.  
 PR 25-SEP-2000: 2000US-0234997.  
 PR 25-SEP-2000: 2000US-0234998.  
 PR 25-SEP-2000: 2000US-0234999.

PR 26-SEP-2000: 2000US-0235484.  
 PR 27-SEP-2000: 2000US-0235834.  
 PR 27-SEP-2000: 2000US-0235836.  
 PR 29-SEP-2000: 2000US-0236327.  
 PR 29-SEP-2000: 2000US-0236367.  
 PR 29-SEP-2000: 2000US-0236368.  
 PR 29-SEP-2000: 2000US-0236369.  
 PR 29-SEP-2000: 2000US-0236370.  
 PR 02-OCT-2000: 2000US-0236802.  
 PR 02-OCT-2000: 2000US-0237037.  
 PR 02-OCT-2000: 2000US-0237038.  
 PR 02-OCT-2000: 2000US-0237039.  
 PR 02-OCT-2000: 2000US-0237040.  
 PR 13-OCT-2000: 2000US-0239937.  
 PR 13-OCT-2000: 2000US-0239938.  
 PR 20-OCT-2000: 2000US-0240960.  
 PR 20-OCT-2000: 2000US-0241221.  
 PR 20-OCT-2000: 2000US-0241785.  
 PR 20-OCT-2000: 2000US-0241786.  
 PR 20-OCT-2000: 2000US-0241787.  
 PR 20-OCT-2000: 2000US-0241808.  
 PR 20-OCT-2000: 2000US-0241809.  
 PR 20-OCT-2000: 2000US-0241826.  
 PR 01-NOV-2000: 2000US-0244617.  
 PR 08-NOV-2000: 2000US-0246474.  
 PR 08-NOV-2000: 2000US-0246475.  
 PR 08-NOV-2000: 2000US-0246476.  
 PR 08-NOV-2000: 2000US-0246477.  
 PR 08-NOV-2000: 2000US-0246478.  
 PR 08-NOV-2000: 2000US-0246523.  
 PR 08-NOV-2000: 2000US-0246524.  
 PR 08-NOV-2000: 2000US-0246525.  
 PR 08-NOV-2000: 2000US-0246526.  
 PR 08-NOV-2000: 2000US-0246527.  
 PR 08-NOV-2000: 2000US-0246528.  
 PR 08-NOV-2000: 2000US-0246532.  
 PR 08-NOV-2000: 2000US-0246609.  
 PR 08-NOV-2000: 2000US-0246610.  
 PR 08-NOV-2000: 2000US-0246611.  
 PR 08-NOV-2000: 2000US-0246613.  
 PR 17-NOV-2000: 2000US-0249207.  
 PR 17-NOV-2000: 2000US-0249208.  
 PR 17-NOV-2000: 2000US-0249209.  
 PR 17-NOV-2000: 2000US-0249210.  
 PR 17-NOV-2000: 2000US-0249211.  
 PR 17-NOV-2000: 2000US-0249212.  
 PR 17-NOV-2000: 2000US-0249213.  
 PR 17-NOV-2000: 2000US-0249214.  
 PR 17-NOV-2000: 2000US-0249215.  
 PR 17-NOV-2000: 2000US-0249216.  
 PR 17-NOV-2000: 2000US-0249217.  
 PR 17-NOV-2000: 2000US-0249218.  
 PR 17-NOV-2000: 2000US-0249244.  
 PR 17-NOV-2000: 2000US-0249245.  
 PR 17-NOV-2000: 2000US-0249246.  
 PR 17-NOV-2000: 2000US-0249247.  
 PR 17-NOV-2000: 2000US-0249248.  
 PR 01-DEC-2000: 2000US-0250160.  
 PR 01-DEC-2000: 2000US-0250391.  
 PR 05-DEC-2000: 2000US-0251030.  
 PR 05-DEC-2000: 2000US-0251988.  
 PR 05-DEC-2000: 2000US-0256719.  
 PR 06-DEC-2000: 2000US-0251479.  
 PR 08-DEC-2000: 2000US-0251856.  
 PR 08-DEC-2000: 2000US-0251868.  
 PR 08-DEC-2000: 2000US-0251869.  
 PR 08-DEC-2000: 2000US-0251989.  
 PR 11-DEC-2000: 2000US-0254097.  
 PR 05-JAN-2001: 2001US-0259678.  
 XX



PA (HUMA-) HUMAN GENOME SCI INC.

PI Rosen CA, Barash SC, Ruben SM;

XX MPI; 2001-488783/53.

DR N-PSDB; AAS26215.

XX New nucleic acid molecules encoding 461 human secreted proteins for  
PT diagnosing, preventing, treating or ameliorating medical conditions and  
PT used as food additives or preservatives.

PS Claim 11; SEQ ID NO 1181; 980bp; English.

XX The invention relates to isolated nucleic acid molecules and their  
CC encoded secreted proteins. The nucleic acids and proteins are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. Antibodies to the proteins can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunoassays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischemia, angioneuromas, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection, and many other  
CC disorders listed in the specification. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence represents a novel secreted protein of the invention.

#### Alignment Scores:

Pred. No.: 2,15e-85 Length: 164  
Score: 813.00 Matches: 153  
Percent Similarity: 100.00% Conservative: 0  
Best Local Similarity: 100.00% Mismatches: 0  
Query Match: 54.60% Indels: 0  
DB: 22 Gaps: 0

US-09-763-902b-10 (1-830) x AAU16228 (1-164)

QY 71 GAATACCCCTGTGAGAGCCCTCCTCATAGAACAGATGGCATCTGTGATTCAG 130  
DB 12 GIIIEPOTRTPRPSPRPSPRPSPRPSPRPSPRPSPRPSPRPSPRPSPRPSP 31  
QY 131 ACCTATGTGATGAGAGCCCTCAGAGCTGCTGAGAGATTTGTCATGCTCTACACACC 190  
DB 32 ThTtYtAlAspGlnAlaCysArgAlaAlaGlnGlnGlnGlnGlnGlnGlnGln 51  
QY 191 ATGATGAAGGCGGCGGCTTGTGCTGCTCCGCTGTACATGGGACACAGCCCTGCTGG 250  
DB 52 MetAspLysAlrGArGArGlnLeuSerArgLeuTyrMetGlyThrAlaThrLeuValTyrp 71  
QY 251 AAGGCAATGCTGCTTTCAGACAAAGATCCTTATAGTACAGTTTGAAGTTCCTCC 310  
DB 72 AAGGlyAsnAlaValSerGlyGlnGlnSerLeuSerGlnGlnGlnGlnGlnGlnGln 91  
QY 311 AGGAGTTCCTCAATCAGCGGTGAGACTGCAGCCTTTCATGATGAGACACACAGC 370  
DB 92 SerGlnPheGlnIleSerValValAspCysGlnProValHisAspGlnAlaThrProSer 111  
QY 371 CAGACACAGGCTCTTGTGTCATCTGTGATGATGATGATGATGATGATGATGATGATGAT 430  
DB 112 GlnThrThrValLeuValValIleCysGlySerValLysPheGlnGlnGlnGlnGlnGln 131  
QY 431 GACTTCAACCAAGACTTCACTGACCGCCAGGCTCCACCCAGCAACACAGTGTGAGAG 490

DB 132 AspheAsnGlnAsnPhelleLeuThrAlaGlnAlaSerProSerAsnThrValTyrpLys 151

QY 491 ATGCAAGTACGCTTCCGCTTCAGAGTGGGCGCAGC 529

DB 152 IleAlaSerAspCysPheArgPheGlnAspThrLaser 164

#### RESULT 2

AA82318  
ID AAY82318 standard; Protein; 140 AA.

XX AAY82318;

DT 19-JUN-2000 (first entry)

XX Human protein transport molecule (PTAM) SEQ ID NO:2.

DE Human; protein transport molecule; PTAM; diagnosis; cytostatic;

XX antiallergic; antidiabetic; immunosuppressant; antihypertensive;

XX dermatological; antianemic; antipruritic; hepatocytic; antipruritic;

XX antineoplastic; antihIV; protein transport regulator; cancer;

XX urticaria; allergy; abnormal vesicle trafficking; asthma;

XX autoimmune haemolytic anaemia.

XX Homo sapiens.

XX WO200012703-A2.

XX 09-MAR-2000.

XX 26-AUG-1999; 99WO-US19616.

XX 27-AUG-1998; 98US-0098206.

XX (INCY-) INCYTE PHARM INC.

XX Tang YT, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ;

XX Gorgone GA, Baughn MR, Patterson C;

XX MPI; 2000-256642/22.

XX N-PSDB; AAA08036.

XX New human protein transport-associated polypeptide and polynucleotide

XX useful for diagnosis, prevention and treatment of cell proliferative

XX and secretory disorders such as, leukemia, cystic fibrosis

XX Claim 1; Page 61-62; 75pp; English.

XX AAA08035 to AAA08042 encode the human protein transport-associated

XX molecules (PTAMs) given in AAY82317 to AAY82324. The PTAMs have

XX cytosolic, antiallergic, antianemic, immunosuppressant,

XX antihypertensive, antidiabetic, antipruritic, antipruritic,

XX antineoplastic, osteopathic, dermatological, antianemic, antipruritic,

XX hepatocytic, antipruritic, antipruritic and antihIV activities, and

XX regulating protein transport. PTAM proteins and antagonists are useful for

XX preventing or treating a disorder associated with decreased or increased

XX expression or activity of PTAM. PTAM polynucleotides are useful for

XX diagnosing conditions associated with PTAM, comprising detecting PTAM by

XX forming a hybridisation complex, preferably after PCR amplifying the

XX biological sample. Diseases prevented, treated or diagnosed include cell

XX proliferative disorders such as cancers, immune disorders, secretory

XX disorders and other conditions associated with abnormal vesicle

XX trafficking, such as allergies, asthma, urticaria and autoimmune

XX haemolytic anaemia. Anti-PTAM antibodies may be used as antagonists, as

XX a targeting or delivery mechanism for bringing pharmaceutical agents

XX into cells or tissues expressing PTAM and for diagnosis of PTAM-related

XX disorders. PTAM, its catalytic or immunogenic fragments are useful for

XX drug screening using libraries of compounds. PTAM polynucleotides are

XX useful for generating hybridisation probes useful in mapping the

XX naturally occurring genomic sequences.

SQ Sequence 140 AA:  
Alignment Scores: 5.22e-76 Length: 140  
Pred. No.: 732.00 Matches: 140  
Score: 100.00% Conservative: 0  
Percent Similarity: 100.00% Mismatches: 0  
Best Local Similarity: 100.00% Indels: 0  
Query Match: 49.168 Gaps: 0  
DB: 21  
US-09-763-902B-10 (1-830) x AAY82318 (1-140)  
QY 110 ATGCGATCGTGGATTTCAGACGCTATGTCATGACGCTCGACGAGCTTCTGAGAGTTT 169  
DB 1 MetAlaSerValAspPheLysThrValAlaSpGlnAlaCysArgAlaAlaGlnGluPhe 20  
QY 170 GTCAATGCTACTACACCCATGATGTAAGCGCGCGCTTCTGCTCCGCTGATACATG 229  
DB 21 ValAsnValTyrTyrThrThrMetAspLysArgArgLeuLeuSerArgLeuTyrMet 40  
QY 230 GGCAGACGACCCCTGGTCTGATGATGCAATGCTGTTTACGACAGAAATCCTTGAGTGAG 289  
DB 41 GylThrAlaThrLeuValTrpAsnGlnAsnAlaValSerGlnGlnGlnSerLeuSerGlu 60  
QY 290 TTTTGTGAATGTGCTTCCAGACGAGTTCCTCAATCAGCGTGAGTACTGCGACGCTGTT 349  
DB 61 PhePheGlnMetLeuProSerSerGluPheGlnIleSerValValAlaSpGlnProVal 80  
QY 350 CATGATGAGCCACACCAAGCCAGACGAGGCTCTTGTCTGATCTGTGATCAGTGAG 409  
DB 81 HisAspGlnAlaThrProSerGlnThrThrValLeuValValIleCysGlnSerValLys 100  
QY 410 TTTGAGGGGACAAACACAGCGGACTTCACAGCAAGAACTTATCTGACGCCAGGCTCA 469  
DB 101 PheGlnGlnLysAlaThrSerLeuAspPheAsnGlnAsnPheIleLeuThrAlaGlnAlaSer 120  
QY 470 CCCAGACACACAGTGTGGAAGATCGCAAGTGAAGTCTGCTCCGCTCCAGAGCTGGGCGCAG 529  
DB 121 ProSerAsnThrValTrpLysIleAlaSerAspCysPheArgPheGlnAspTrpAlaSer 140  
RESULT 3  
AAB95379  
ID AAB95379 standard; Protein; 197 AA.  
XX AC AAB95379;  
XX 26-JUN-2001 (first entry)  
XX Human protein sequence SEQ ID NO:17707.  
XX DE Human protein sequence SEQ ID NO:17707.  
XX KM Human; primer; detection; diagnosis; antisense therapy; gene therapy.  
XX OS Homo sapiens.  
XX PN EP1074617-A2.  
XX PD 07-FEB-2001.  
XX PF 28-JUL-2000; 2000EP-0116126.  
XX PR 29-JUL-1999; 99JP-0248036.  
XX PR 27-AUG-1999; 99JP-0300253.  
XX PR 11-JAN-2000; 2000JP-0118776.  
XX PR 02-MAY-2000; 2000JP-0183767.  
XX PR 09-JUN-2000; 2000JP-0241899.  
XX PA (HELI-) HELIX RES INST.  
XX PI Ota T, Isogai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;  
XX Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;  
XX WPI; 2001-318749/34.

PT Primer sets for synthesizing polynucleotides, particularly the 5602  
PT full-length cDNAs defined in the specification, and for the detection  
PT and/or diagnosis of the abnormality of the proteins encoded by the  
PT full-length cDNAs -  
Claim 8; SEQ ID 17707; 2537pp + CD ROW; English.  
CC The present invention describes primer sets for synthesizing 5602  
CC full-length cDNAs defined in the specification. Where a primer set  
CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary  
CC to the complementary strand of a polynucleotide which comprises one of  
CC the 5602 nucleotide sequences defined in the specification, where the  
CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination  
CC of an oligonucleotide comprising a sequence complementary to the  
CC complementary strand of a polynucleotide which comprises a 5'-end  
CC sequence and an oligonucleotide comprising a sequence complementary to a  
CC polynucleotide which comprises a 3'-end sequence, where the  
CC oligonucleotide comprises at least 15 nucleotides and the combination of  
CC the 5'-end sequence/3'-end sequence is selected from those defined in  
CC the specification. The primer sets can be used in antisense therapy and  
CC in gene therapy. The primers are useful for synthesizing polynucleotides,  
CC particularly full-length cDNAs. The primers are also useful for the  
CC detection and/or diagnosis of the abnormality of the proteins encoded by  
CC the full-length cDNAs. The primers allow obtaining of the full-length  
CC cDNAs easily without any specialized methods. AAH03166 to AAH13628 and  
CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to  
CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632  
CC represent oligonucleotides, all of which are used in the exemplification  
CC of the present invention.  
XX  
XX  
SQ Sequence 197 AA:  
Alignment Scores: 8.82e-59 Length: 197  
Pred. No.: 584.50 Matches: 106  
Score: 90.85% Conservative: 23  
Percent Similarity: 74.65% Mismatches: 12  
Best Local Similarity: 39.25% Indels: 1  
Query Match: 22 Gaps: 1  
DB: 22  
US-09-763-902B-10 (1-830) x AAB95379 (1-197)  
QY 107 GAGATGGCA--TCTGTGATTTCAAGACCTATGTGATCAGGCTCGACAGCTGCTGAG 163  
DB 55 GlnMetAlaThrSerLeuAspPheLysThrTyrValAlaSpGlnAlaCysArgAlaAlaGln 74  
QY 164 GAGTTGTGAATGTCTACTACACACCATGATGATACGGCGGCGTGTCTGCTCCGCGC 223  
DB 75 GluPheValAlaSerIleTyrTyrGlnThrMetAspLysArgArgAlaLeuThrArgLeu 94  
QY 224 TGCATGGGACACGCGCCCTGGTCTGATGGAATGGAATGCTTTCAGACAGCAAGATCCTG 283  
DB 95 TyrLeuAspLysAlaThrLeuIleTrpAsnGlnAlaValSerIleLeuAspAlaLeu 114  
QY 284 ACTGAGTTTGTGAATGTGCTTCCAGAGTTCCTCAATCAGCTGAGTACTGCCAG 343  
DB 115 AsnAsnPhePheAspThrLeuProSerSerGluPheGlnValAlaSerLeuAspCysGln 134  
QY 344 CCTGTTCATGATGAAACCAACCAACCAACGACGCTCTGTTCTCATCTGTGATCA 403  
DB 135 ProValHisGlnGlnAlaThrGlnSerGlnThrThrValLeuValAlaThrSerGlnTyr 154  
QY 404 GTGAAGTTGAGGGGACAAACACAGCGGACTTCACCAAGTTCATCTGACCGCCAG 463  
DB 155 ValLysPheAspLysAlaThrLeuIleTrpAsnGlnAlaValSerIleLeuAspAlaLeu 174  
QY 464 GCCTACCCAGACACAGTGTGGAAGATCGCAAGTGAAGTCTGCTCCGCTCCAGAGCTGG 523  
DB 175 SerThrProAsnAsnThrValTrpLysIleAlaSerAspCysPheArgPheGlnAspTrp 194  
QY 524 GCCAGC 529  
DB 195 SerSer 196

RESULT 4  
AAV82323  
ID AAV82323 standard; Protein: 142 AA.  
AC AAV82323;  
DT 19-JUN-2000 (first entry)  
XX  
XX Human protein transport molecule (PRAM) SEQ ID NO:7.  
DE  
XX Human; protein transport molecule; PRAM; diagnosis; cytostatic;  
XX antiallergic; antidiabetic; immunosuppressant; antiarteriosclerotic;  
XX antiallergic; antidiabetic; antileptemic; antirheumatic; osteopathic;  
XX dermatological; antianemic; antipruritic; hepatotropic; antipruritic;  
XX antinflammatory; antihiv; protein transport; regulator; cancer;  
XX immune disorder; cell proliferative disorder; secretory disorder;  
XX urticaria; allergy; abnormal vesicle trafficking; asthma;  
XX autoimmune haemolytic anaemia.  
XX  
XX Homo sapiens.  
XX  
XX WC0200012703-A2.  
XX  
XX 09-MAR-2000.  
XX  
XX 26-AUG-1999; 99MO-US19616.  
XX  
XX 27-AUG-1998; 98US-0098206.  
XX  
XX (INCY-) INCYTE PHARM INC.  
XX  
XX Tang YT, Lal P, Bandman O, Yue H, Corley NC, Guegler KJ;  
XX Gorgone GA, Baughn MR, Patterson C;  
XX  
XX WPI: 2000-256642/22.  
XX  
XX DR N-PSDB: AAA08041.  
XX  
XX New human protein transport-associated polypeptide and polynucleotide  
XX useful for diagnosis, prevention and treatment of cell proliferative  
XX and secretory disorders such as leukemia, cystic fibrosis  
XX  
XX  
PS Claim 1; Page 68; 75pp; English.  
XX  
XX AAA08035 to AAA08042 encode the human protein transport-associated  
XX molecules (PRAMs) given in AAV82317 to AAV82324. The PRAMs have  
XX cytoskeletal, antiallergic, antidiabetic, immunosuppressant,  
XX antirheumatic, antileptemic, antipruritic, antipruritic,  
XX antineurotic, osteopathic, dermatological, antianemic, antipruritic,  
XX hepatotropic, antipruritic, antinflammatory and antihiv activities, and  
XX regulate protein transport. PRAM proteins and antagonists are useful for  
XX preventing or treating a disorder associated with decreased or increased  
XX expression or activity of PRAM. PRAM polynucleotides are useful for  
XX diagnosing conditions associated with PRAM, comprising detecting PRAM by  
XX forming a hybridisation complex, preferably after PCR amplifying the  
XX biological sample. Diseases prevented, treated or diagnosed include cell  
XX proliferative disorders such as cancers, immune disorders, secretory  
XX disorders and other conditions associated with abnormal vesicle  
XX trafficking, such as allergies, asthma, urticaria and autoimmune  
XX haemolytic anaemia. Anti-PRAM antibodies may be used as antagonists, as  
XX a targeting or delivery mechanism for bringing pharmaceutical agents,  
XX into cells or tissues expressing PRAM and for diagnosis of PRAM-related  
XX disorders. PRAM, its catalytic or immunogenic fragments are useful for  
XX drug screening using libraries of compounds. PRAM polynucleotides are  
XX useful for generating hybridisation probes useful in mapping the  
XX naturally occurring genomic sequences.  
XX  
XX Sequence 142 AA:  
SO  
Alignment Scores: 8.57e-59 Length: 142  
Pred. No.: 584.00 Matches: 104  
Score: 91.30% Conservative: 22  
Percent Similarity: 91.30%

Best Local Similarity: 75.36% Mismatches: 12  
Query Match: 39.22% Indels: 0  
DB: 21 Gaps: 0  
US-09-763-902b-10 (1-830) x AAV82323 (1-142)  
QY 116 TCCTGATTTTCAAGACCTATGTTGATTCAGGCTGACAGAGCTGCTGAGACTTTGTCAAT 175  
DB 4 SerLeuAspPheLysThrTyValAspGlnAlaLysArgAlaLysGluGluPheValAsn 23  
QY 176 GTCTACTACACACACCATGATTAAGCGCGCGCTGCTGCTGCCCTGCTACATGGGCACA 235  
DB 24 ILCTYTYGluThrMetAspLysArgArgAlaLeuThrAlaGluTYrLeuAspLys 43  
QY 236 GCCACCGCTGGCTGGATGCAATGCGTTTTCAGACACAGAACTGAGAGATTGTTT 295  
DB 44 AlathrLeuThrPheAsnGlyAsnAlaValSerGlyLeuAspAlaLeuAsnAspPhe 63  
QY 296 GAATGTGCTTCCAGCGAGTTCCAAATCAGCTGAGTCCAGCTGATCATGAT 355  
DB 64 AspThrLeuProSerSerGluPheGlnValAsnMetLeuAspGlyGlnProValHisGlu 83  
QY 356 GAAGCCACACACACACACAGCCAGCGCTGCTGCTGCTGATGATGATGATGATGATGATGAT 415  
DB 84 GlnAlaThrGlnSerGlnThrThrValLeuValAlaThrSerGlyThrValLysPheAsp 103  
QY 416 GGGAAACAACAACAGCGAGCTTCAACCACTTCACCGAGCGCGCGCTGACCCAGC 475  
DB 104 GlyAsnLysGlnHisPhePheAsnGlnAsnPheLeuThrAlaGlnSerThrProAsn 123  
QY 476 AACACAGTGTGGAAGATCGCAAGTACTGCTCCGCTTCCAGAGCTGGCCAGC 529  
DB 124 AsnThrValThrPheLysIleAlaSerAspLysPheArgPheGlnAsnThrProSer 141  
RESULT 5  
ABR04459  
ID ABR04459 standard; Protein: 142 AA.  
XX  
XX ABR04459;  
XX  
XX 04-MAR-2002 (first entry)  
XX  
XX  
DE Human NTF2 associated protein 16.  
XX  
XX Human; NTF2 associated protein 16; cancer; HIV infection; gene therapy.  
XX  
XX Homo sapiens.  
XX  
XX CN133332-A.  
XX  
XX 19-SEP-2001.  
XX  
XX 15-MAR-2000; 2000CN-0114918.  
XX  
XX 15-MAR-2000; 2000CN-0114918.  
XX  
XX (BODE-) BODE GENE DEV CO LTD SHANGHAI.  
XX  
XX Mao Y, Xie Y;  
XX  
XX WPI: 2002-042199/06.  
XX  
XX DR N-PSDB: ABR05754.  
XX  
XX Polypeptide-human NTF2 associated protein 16 and polynucleotide  
XX encoding it  
XX  
PS Claim 1; Page 26(Disclosure); 32pp; Chinese.  
XX  
XX The present invention provides the protein and coding sequences of human  
XX NTF2 associated protein 16. The sequences can be used in the treatment of  
XX cancer and HIV infection, as well as other diseases. The present sequence  
XX is the protein of the invention.



PR 13-OCT-2000; 2000US-0239937.  
PR 20-OCT-2000; 2000US-0240960.  
PR 20-OCT-2000; 2000US-0241221.  
PR 20-OCT-2000; 2000US-0241785.  
PR 20-OCT-2000; 2000US-0241786.  
PR 20-OCT-2000; 2000US-0241787.  
PR 20-OCT-2000; 2000US-0241808.  
PR 20-OCT-2000; 2000US-0241809.  
PR 20-OCT-2000; 2000US-0241826.  
PR 01-NOV-2000; 2000US-0246417.  
PR 08-NOV-2000; 2000US-0246474.  
PR 08-NOV-2000; 2000US-0246475.  
PR 08-NOV-2000; 2000US-0246476.  
PR 08-NOV-2000; 2000US-0246477.  
PR 08-NOV-2000; 2000US-0246478.  
PR 08-NOV-2000; 2000US-0246523.  
PR 08-NOV-2000; 2000US-0246524.  
PR 08-NOV-2000; 2000US-0246525.  
PR 08-NOV-2000; 2000US-0246526.  
PR 08-NOV-2000; 2000US-0246527.  
PR 08-NOV-2000; 2000US-0246528.  
PR 08-NOV-2000; 2000US-0246532.  
PR 08-NOV-2000; 2000US-0246509.  
PR 08-NOV-2000; 2000US-0246610.  
PR 08-NOV-2000; 2000US-0246611.  
PR 08-NOV-2000; 2000US-0246613.  
PR 17-NOV-2000; 2000US-0249207.  
PR 17-NOV-2000; 2000US-0249208.  
PR 17-NOV-2000; 2000US-0249209.  
PR 17-NOV-2000; 2000US-0249210.  
PR 17-NOV-2000; 2000US-0249211.  
PR 17-NOV-2000; 2000US-0249212.  
PR 17-NOV-2000; 2000US-0249213.  
PR 17-NOV-2000; 2000US-0249214.  
PR 17-NOV-2000; 2000US-0249215.  
PR 17-NOV-2000; 2000US-0249216.  
PR 17-NOV-2000; 2000US-0249217.  
PR 17-NOV-2000; 2000US-0249218.  
PR 17-NOV-2000; 2000US-0249219.  
PR 17-NOV-2000; 2000US-0249244.  
PR 17-NOV-2000; 2000US-0249245.  
PR 17-NOV-2000; 2000US-0249246.  
PR 17-NOV-2000; 2000US-0249255.  
PR 17-NOV-2000; 2000US-0249257.  
PR 17-NOV-2000; 2000US-0249259.  
PR 17-NOV-2000; 2000US-0249300.  
PR 01-DEC-2000; 2000US-0250160.  
PR 01-DEC-2000; 2000US-0250391.  
PR 05-DEC-2000; 2000US-0251030.  
PR 05-DEC-2000; 2000US-0251988.  
PR 05-DEC-2000; 2000US-0256719.  
PR 06-DEC-2000; 2000US-0251479.  
PR 08-DEC-2000; 2000US-0251856.  
PR 08-DEC-2000; 2000US-0251868.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251869.  
PR 08-DEC-2000; 2000US-0251990.  
PR 11-DEC-2000; 2000US-0254097.  
PR 05-JAN-2001; 2001US-0259678.  
XX  
XX (HUMA-) HUMAN GENOME SCI INC.  
XX  
XX Rosen CA, Barash SC, Ruben SM;  
XX  
XX WPI: 2001-488783/53.  
XX  
XX N-PSDB; AAS26214.  
XX  
XX New nucleic acid molecules encoding 461 human secreted proteins for  
XX PT diagnosing, preventing, treating or ameliorating medical conditions and  
XX PT used as food additives or preservatives -  
XX  
XX Claim 11; SEQ ID NO 1180; 980pp; English.  
XX  
XX The invention relates to isolated nucleic acid molecules and their

CC encoded secreted proteins. The nucleic acids and proteins are used to  
CC prevent, treat or ameliorate a medical condition in e.g. humans, mice,  
CC rabbits, goats, horses, cats, dogs, chickens or sheep. They  
CC are also used in diagnosing a pathological condition or susceptibility  
CC to a pathological condition. Antibodies to the proteins can also  
CC be used in alleviating symptoms associated with the disorders and in  
CC diagnostic immunoassays e.g. radioimmunoassays or enzyme linked  
CC immunoassay assays (ELISA). Disorders which are diagnosed or treated  
CC include autoimmune diseases e.g. Rheumatoid arthritis,  
CC hyperproliferative disorders e.g. neoplasms of the breast or liver,  
CC cardiovascular disorders e.g. cardiac arrest, cerebrovascular disorders  
CC e.g. cerebral ischemia, angiodenesis, nervous system disorders e.g.  
CC Alzheimer's disease, infections caused by bacteria, viruses and fungi  
CC and ocular disorders e.g. corneal infection, and many other  
CC disorders listed in the specification. The polypeptides can also  
CC be used to aid wound healing and epithelial cell proliferation, to  
CC prevent skin aging due to sunburn, to maintain organs before  
CC transplantation, for supporting cell culture of primary tissues, to  
CC regenerate tissues and in chemotaxis. The polypeptides can also be used  
CC as a food additive or preservative to increase or decrease storage  
CC capabilities, fat content, lipid, protein, carbohydrate, vitamins,  
CC minerals, cofactors and other nutritional components. The present  
CC sequence represents a novel secreted protein of the invention.

## Alignment Scores:

Pred. No.:	3,496-58	Length:	162
Score:	579.00	Matches:	103
Percent Similarity:	91.18%	Conservative:	21
Best Local Similarity:	75.74%	Mismatches:	12
Query Match:	38.89%	Indels:	0
DB:	22	Gaps:	0

US-09-763-902b-10 (1-830) x AAUI6227 (1-162)

QY	122	GATTTCAAGACCTATGTGATCAGAGCCCTGACAGAGTGTGATGATTTGTAATGTCTAC	181
DB	26	AspPheLysTrpTrpValValSpGlnAlaCysArgAlaAlaGluGluPheValAlaSer	45
QY	182	TACACACCATGATGATTAAGCGCGCGTGTGTCGCCCTGATCATGGGACAGACACC	241
DB	46	TyrGlnTrpMetAspLysArgArgArgAlaLeuTrpArgLeuTrpLysAspLysAlaThr	65
QY	242	CTGCTCTGGAATGGCATGCTGTTTCAAGACAGACAGATCTGAGTGTGTTTGAATG	301
DB	66	LeuIleTrpAsnGlnAlaValSerGlyLeuAspAlaLeuAsnAspPheSer	85
QY	302	TTCCTTCACGAGCTTCCAAATCAGCGCTGATGATCCAGCCCTGATCAGAGACCC	361
DB	86	LeuProSerSerGluPheGlnValAlaMetLeuAspCysGlnProValHisGlnIleAla	105
QY	362	ACACCAAGCAGACAGCAGCGCTGTTGTCATCTGTGATGATAGTGAATTTGAGGAGAC	421
DB	106	ThrGlnSerGlnTrpThrValLeuValAlaThrSerGlyTrpValLysPheAspLysn	125
QY	422	AAACAACGCGACTTCAACAGCAACTTCACTGACCGCCAGCGCCCTCAGCAACACA	481
DB	126	LysGlnHisPhePheAsnGlnAsnPheLeuLeuTrpAlaGlnSerThrProAsnAsnThr	145
QY	482	GTCGTGAAGATGCGAGTGAAGTGTCTCCGCTTCCAGAGATGGCCAGC	529
DB	146	ValTrpLysIleAlaSerAspCysPheArgPheGlnAspTrpSerSer	161

## RESULT 7

ABB65017 standard; Protein: 133 AA.

ABB65017;

26-MAR-2002 (first entry)

XX Drosophila melanogaster polypeptide SEQ ID NO 21843.  
XX DE  
XX Drosophila; developmental biology; cell signalling; insecticide;  
KW

KM pharmaceutical.  
XX Drosophila melanogaster.  
XX  
XX WO200171042-A2.  
XX  
XX 27-SEP-2001.  
XX  
XX 23-MAR-2001; 2001WO-US092331.  
XX  
XX 23-MAR-2000; 2000US-191637P.  
XX 11-JUL-2000; 2000US-0614150.  
XX  
XX (PEKE ) PE CORP NY.  
XX  
XX Venter JC, Adams M, Li PND, Myers EW;  
XX  
XX WPI; 2001-656860/75.  
XX  
XX N-PSDB; ABL09120.  
XX  
XX  
XX New isolated nucleic acid detection reagent for detecting 1000 or more  
XX genes from Drosophila and for elucidating cell signalling and cell-cell  
XX interactions  
XX  
XX Disclosure; SEQ ID NO 21843; 21bp + Sequence Listing; English.  
XX  
XX  
XX The invention relates to an isolated nucleic acid detection reagent  
XX capable of detecting 1000 or more genes from Drosophila. The invention is  
XX useful in developmental biology and in elucidating cell signalling and  
XX cell-cell interactions in higher eukaryotes for the development of  
XX insecticides, therapeutics and pharmaceutical drugs. The invention  
XX discloses genomic DNA sequences (AB101840-AB16175), expressed DNA  
XX sequences (AB101840-AB16175) and the encoded proteins  
XX (AB57737-AB572072).  
XX  
XX The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic format directly from WIPO  
XX at ftp.wipo.int/pub/published\_pat\_sequences.  
XX  
XX  
XX Sequence 133 AA:  
XX  
XX  
XX Alignment Scores:  
XX  
XX Pred. No.: 1,04e-26 Length: 133  
XX Score: 308.00 Matches: 55  
XX Percent Similarity: 61.65% Conservative: 27  
XX Best Local Similarity: 41.35% Mismatches: 47  
XX Query Match: 20.69% Indels: 4  
XX DB: 22 Gaps: 1  
XX  
XX US-09-763-902B-10 (1-830) x ABB65017 (1-133)  
XX  
XX  
XX 122 GATTTCAGACCTATGTGATCGAGCGCTGAGAGTTCATGCTC 181  
XX  
XX 4 AspleubylsalalysvalgluserCysAlaArgThrAlasphrPheThrArgLeuTy 23  
XX  
XX 182 TACACACCATGTGATAGCGCGCGCTTCCTCCCGCTGACATGGGACACAGCAC 241  
XX  
XX 24 TyAlaservalaspsnAArgAArgInLleGlyArgLeuTyTyreusApsanAlaThr 43  
XX  
XX 242 CTGCTGTGAGATGCAATGCTGTTTCAGACAGAAATCCCTGAGTGGATTTTGAATG 301  
XX  
XX 44 LeuSerTyrpsnLysnInLleGlyArgLleGlyMetIleGlyLeuSerTyrPheGln 63  
XX  
XX 302 TTGCTTCACGAGTTCACAAATCAGCGGTGACTGCGACCTGTCTATGATGAAC 361  
XX  
XX 64 LeuPyrSerSerTyrnhtsGlnLeuAAsnThrLeuApsAlaGlnProIleValAaspGlnAla 83  
XX  
XX 362 ACACCAACGACCAACGAGTCTGTGTCATCTGTGATCAGTGAAGTTGAGGGAGAC 421  
XX  
XX 84 ValSerAsnGlnLeuAlaTyrLeuIleMetAlaSerGlySerValLysPheAlaAspGln 103  
XX  
XX 422 AAACACAGGACCTTCACACGAATCTCATCTTACAGCGCCAGGCTTACCACACACA 481  
XX  
XX 104 GlnLeuArgLysPheGlnGlnThrPheIleValThrAlaGlu-----AsnAsp 119

QY 482 GTGTGAGATGCGCAAGTACTGCTTCGCCCTTCACAGAC 520  
DB 120 LysTrpLysValValSerAspCysTyrArgTyrMetGln 132  
RESULT 8  
ID AAG12073 standard; Protein; 125 AA.  
AC AAG12073;  
XX  
XX 17-OCT-2000 (first entry)  
XX  
XX Arabidopsis thaliana protein fragment SEQ ID NO: 11042.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
XX hybridisation assay; genetic mapping; gene expression control; promoter;  
XX termination sequence.  
XX  
XX Arabidopsis thaliana.  
XX  
XX  
XX EPI03405-A2.  
XX  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
XX  
XX  
XX 25-FEB-1999; 99US-0121825.  
XX 05-MAR-1999; 99US-0123180.  
XX 09-MAR-1999; 99US-0123548.  
XX 23-MAR-1999; 99US-0125788.  
XX 25-MAR-1999; 99US-0126264.  
XX 29-MAR-1999; 99US-0126785.  
XX 01-APR-1999; 99US-0127462.  
XX 06-APR-1999; 99US-0128234.  
XX 08-APR-1999; 99US-0128714.  
XX 16-APR-1999; 99US-0129845.  
XX 19-APR-1999; 99US-0130077.  
XX 21-APR-1999; 99US-0130449.  
XX 23-APR-1999; 99US-0130510.  
XX 23-APR-1999; 99US-0130891.  
XX 28-APR-1999; 99US-0131449.  
XX 30-APR-1999; 99US-0132048.  
XX 30-APR-1999; 99US-0132407.  
XX 04-MAY-1999; 99US-0132484.  
XX 05-MAY-1999; 99US-0132485.  
XX 06-MAY-1999; 99US-0132486.  
XX 06-MAY-1999; 99US-0132487.  
XX 07-MAY-1999; 99US-0132863.  
XX 11-MAY-1999; 99US-0134256.  
XX 14-MAY-1999; 99US-0134258.  
XX 14-MAY-1999; 99US-0134219.  
XX 14-MAY-1999; 99US-0134221.  
XX 14-MAY-1999; 99US-0134370.  
XX 18-MAY-1999; 99US-0134768.  
XX 19-MAY-1999; 99US-0134941.  
XX 20-MAY-1999; 99US-0135124.  
XX 21-MAY-1999; 99US-0135353.  
XX 24-MAY-1999; 99US-0135629.  
XX 25-MAY-1999; 99US-0136021.  
XX 27-MAY-1999; 99US-0136392.  
XX 28-MAY-1999; 99US-013782.  
XX 01-JUN-1999; 99US-0137222.  
XX 03-JUN-1999; 99US-0137528.  
XX 04-JUN-1999; 99US-0137502.  
XX 07-JUN-1999; 99US-0137724.  
XX 08-JUN-1999; 99US-0138094.  
XX 10-JUN-1999; 99US-0138540.  
XX 10-JUN-1999; 99US-0138847.  
XX 14-JUN-1999; 99US-0139119.  
XX 16-JUN-1999; 99US-0139452.  
XX 16-JUN-1999; 99US-0139453.  
XX 17-JUN-1999; 99US-0139492.







PR	23-JUL-1999	9905-0145119
PR	23-JUL-1999	9905-0145121
PR	23-JUL-1999	9905-0145216
PR	26-JUL-1999	9905-0145224
PR	26-JUL-1999	9905-0145276
PR	27-JUL-1999	9905-0145911
PR	27-JUL-1999	9905-0145918
PR	27-JUL-1999	9905-0145918
PR	28-JUL-1999	9905-0145951
PR	28-JUL-1999	9905-0145951
PR	02-AUG-1999	9905-0146386
PR	02-AUG-1999	9905-0146388
PR	03-AUG-1999	9905-0147038
PR	04-AUG-1999	9905-0147204
PR	04-AUG-1999	9905-0147302
PR	05-AUG-1999	9905-0147192
PR	05-AUG-1999	9905-0147260
PR	06-AUG-1999	9905-0147303
PR	06-AUG-1999	9905-0147416
PR	09-AUG-1999	9905-0147493
PR	09-AUG-1999	9905-0147935
PR	10-AUG-1999	9905-0148171
PR	11-AUG-1999	9905-0148341
PR	12-AUG-1999	9905-0148341
PR	13-AUG-1999	9905-0148565
PR	13-AUG-1999	9905-0148684
PR	16-AUG-1999	9905-0149368
PR	17-AUG-1999	9905-0149175
PR	18-AUG-1999	9905-0149426
PR	20-AUG-1999	9905-0149722
PR	20-AUG-1999	9905-0149723
PR	20-AUG-1999	9905-0149923
PR	20-AUG-1999	9905-0149923
PR	23-AUG-1999	9905-0149902
PR	23-AUG-1999	9905-0149930
PR	25-AUG-1999	9905-0150566
PR	26-AUG-1999	9905-0150584
PR	27-AUG-1999	9905-0151065
PR	27-AUG-1999	9905-0151066
PR	27-AUG-1999	9905-0151080
PR	30-AUG-1999	9905-0151303
PR	31-AUG-1999	9905-0151438
PR	01-SEP-1999	9905-0151930
PR	07-SEP-1999	9905-0152363
PR	10-SEP-1999	9905-0153070
PR	13-SEP-1999	9905-0153758
PR	15-SEP-1999	9905-0154018
PR	16-SEP-1999	9905-0154039
PR	20-SEP-1999	9905-0154773
PR	22-SEP-1999	9905-0155133
PR	23-SEP-1999	9905-0155486
PR	24-SEP-1999	9905-0155659
PR	26-SEP-1999	9905-0156458
PR	28-SEP-1999	9905-0156596
PR	04-OCT-1999	9905-0157117
PR	05-OCT-1999	9905-0157733
PR	06-OCT-1999	9905-0157865
PR	07-OCT-1999	9905-0158029
PR	08-OCT-1999	9905-0158232
PR	12-OCT-1999	9905-0158369
PR	13-OCT-1999	9905-0158293
PR	13-OCT-1999	9905-0158294
PR	14-OCT-1999	9905-0158294
PR	14-OCT-1999	9905-0158637
PR	14-OCT-1999	9905-0158637
PR	18-OCT-1999	9905-0158584
PR	21-OCT-1999	9905-0160741
PR	21-OCT-1999	9905-0160747
PR	21-OCT-1999	9905-0160768
PR	21-OCT-1999	9905-0160770
PR	21-OCT-1999	9905-0160814

[illegible]

**Alignment Scores:**

Pred. No.:	120-05	Length:	15
Score:	120.00	Matches:	43
Percent Similarity:	46.31%	Conservative:	26
Best Local Similarity:	28.86%	Mismatches:	58
Query Match:	8.06%	Indels:	22
DB:	21	Gaps:	7

US-09-763-902B-10 (1-830) x AAG05145 (1-153)

86 AGCCCTCCTTCATAGAACGAGATGGCATCTGTGGATTTCAAGACCCTATGTGGATCAG 145

Db 12 SerSerAlaSerLeuSerLeuLysMetSerGlnMetasp-----ProAspAla 27

2Y 146 GCCCTGCAGAGCTGCTGAGGAGTTGTCAATGCTACTACACCACCAATGGATAAGCGCGG 203

20 valisei lysala ----- pnevalovunsiyilyisei in pnevpi nstiaiy 44

[illegible]

45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000 1001 1002 1003 1004 1005 1006 1007 1008 1009 1010 1011 1012 1013 1014 1015 1016 1017 1018 1019 1020 1021 1022 1023 1024 1025 1026 1027 1028 1029 1030 1031 1032 1033 1034 1035 1036 1037 1038 1039 1040 1041 1042 1043 1044 1045 1046 1047 1048 1049 1050 1051 1052 1053 1054 1055 1056 1057 1058 1059 1060 1061 1062 1063 1064 1065

[illegible][illegible]

2  
 3  
 4  
 5  
 6  
 7  
 8  
 9  
 10  
 11  
 12  
 13  
 14  
 15  
 16  
 17  
 18  
 19  
 20  
 21  
 22  
 23  
 24  
 25  
 26  
 27  
 28  
 29  
 30  
 31  
 32  
 33  
 34  
 35  
 36  
 37  
 38  
 39  
 40  
 41  
 42  
 43  
 44  
 45  
 46  
 47  
 48  
 49  
 50  
 51  
 52  
 53  
 54  
 55  
 56  
 57  
 58  
 59  
 60  
 61  
 62  
 63  
 64  
 65  
 66  
 67  
 68  
 69  
 70  
 71  
 72  
 73  
 74  
 75  
 76  
 77  
 78  
 79  
 80  
 81  
 82  
 83  
 84  
 85  
 86  
 87  
 88  
 89  
 90  
 91  
 92  
 93  
 94  
 95  
 96  
 97  
 98  
 99  
 100  
 101  
 102  
 103  
 104  
 105  
 106  
 107  
 108  
 109  
 110  
 111  
 112  
 113  
 114  
 115  
 116  
 117  
 118  
 119  
 120  
 121  
 122  
 123  
 124  
 125  
 126  
 127  
 128  
 129  
 130  
 131  
 132  
 133  
 134  
 135  
 136  
 137  
 138  
 139  
 140  
 141  
 142  
 143  
 144  
 145  
 146  
 147  
 148  
 149  
 150  
 151  
 152  
 153  
 154  
 155  
 156  
 157  
 158  
 159  
 160  
 161  
 162  
 163  
 164  
 165  
 166  
 167  
 168  
 169  
 170  
 171  
 172  
 173  
 174  
 175  
 176  
 177  
 178  
 179  
 180  
 181  
 182  
 183  
 184  
 185  
 186  
 187  
 188  
 189  
 190  
 191  
 192  
 193  
 194  
 195  
 196  
 197  
 198  
 199  
 200  
 201  
 202  
 203  
 204  
 205  
 206  
 207  
 208  
 209  
 210  
 211  
 212  
 213  
 214  
 215  
 216  
 217  
 218  
 219  
 220  
 221  
 222  
 223  
 224  
 225  
 226  
 227  
 228  
 229  
 230  
 231  
 232  
 233  
 234  
 235  
 236  
 237  
 238  
 239  
 240  
 241  
 242  
 243  
 244  
 245  
 246  
 247  
 248  
 249  
 250  
 251  
 252  
 253  
 254  
 255  
 256  
 257  
 258  
 259  
 260  
 261  
 262  
 263  
 264  
 265  
 266  
 267  
 268  
 269  
 270  
 271  
 272  
 273  
 274  
 275  
 276  
 277  
 278  
 279  
 280  
 281  
 282  
 283  
 284  
 285  
 286  
 287  
 288  
 289  
 290  
 291  
 292  
 293  
 294  
 295  
 296  
 297  
 298  
 299  
 300  
 301  
 302  
 303  
 304  
 305  
 306  
 307  
 308  
 309  
 310  
 311  
 312  
 313  
 314  
 315  
 316  
 317  
 318  
 319  
 320  
 321  
 322  
 323  
 324  
 325  
 326  
 327  
 328  
 329  
 330  
 331  
 332  
 333  
 334  
 335  
 336  
 337  
 338  
 339  
 340  
 341  
 342  
 343  
 344  
 345  
 346  
 347  
 348  
 349  
 350  
 351  
 352  
 353  
 354  
 355  
 356  
 357  
 358  
 359  
 360  
 361  
 362  
 363  
 364  
 365  
 366  
 367  
 368  
 369  
 370  
 371  
 372  
 373  
 374  
 375  
 376  
 377  
 378  
 379  
 380  
 381  
 382  
 383  
 384  
 385  
 386  
 387  
 388  
 389  
 390  
 391  
 392  
 393  
 394  
 395  
 396  
 397  
 398  
 399  
 400  
 401  
 402  
 403  
 404  
 405  
 406  
 407  
 408  
 409  
 410  
 411  
 412  
 413  
 414  
 415  
 416  
 417  
 418  
 419  
 420  
 421  
 422  
 423  
 424  
 425  
 426  
 427  
 428  
 429  
 430  
 431  
 432  
 433  
 434  
 435  
 436  
 437  
 438  
 439  
 440  
 441  
 442  
 443  
 444  
 445  
 446  
 447  
 448  
 449  
 450  
 451  
 452  
 453  
 454  
 455  
 456  
 457  
 458  
 459  
 460  
 461  
 462  
 463  
 464  
 465  
 466  
 467  
 468  
 469  
 470  
 471  
 472  
 473  
 474  
 475  
 476  
 477  
 478  
 479  
 480  
 481  
 482  
 483  
 484  
 485  
 486  
 487  
 488  
 489  
 490  
 491  
 492  
 493  
 494  
 495  
 496  
 497  
 498  
 499  
 500  
 501  
 502  
 503  
 504  
 505  
 506  
 507  
 508  
 509  
 510  
 511  
 512  
 513  
 514  
 515  
 516  
 517  
 518  
 519  
 520  
 521  
 522  
 523  
 524  
 525  
 526

[illegible][illegible]

007 437 AACCACCACTTCCATCCTG-----ACCCCCCACCCCACCAACCAACAAGTCGGAAG 498

```

pb          ::::||| ||| ||| ||| |
120 Src]MetPhen]LeuMetBroThyProG]nG]vSer-----pheTyr 134

```

491 ATGGCAAGTGACTGCTTCCGCTTCAG 517

Db 135 ValPheAsnAspIlePheSerTrpArg 143.

RESULT 10

ABB67404  
ID ABB67404 standard; Protein; 688 AA.

AC . ABB67404;

DT 26-MAR-2002 (first entry)

DE Drosophila melanogaster polypeptide SEQ ID NO 29004.

KW *Drosophila*; developmental biology; cell signalling; insecticide;

[illegible]

XX  
PN  
WC300171043-A3



OY 335 GATCCGACCTGTCATGATGACCAACCAACGACGAGCTCTCTGTCATC 394  
||| |||  
Db AspaIaGlnAlaThLeuGlyAsnGly-----ValValAlaGlnVal 92  
OY 395 TGTGATGATGACCTGTTGAGGGAACAACGAGCTTCACAGACTTCATCCG 454  
||| |||  
Db 93 ThrGlyGlnLeuSerAsnAspGlyGlnPrometargArgPheThGlnThrPheValLeu 112  
OY 455 ACCGCCGACGCTCCACCAACGACAGTGTGAGATGCGAAGTACTGCTCCGCTTC 514  
|||||  
Db 113 AlaAlaGln-----SerProLys-----TyrTyrValHisAsnAspIlePheArgTyr 129  
OY 515 CAGGAC 520  
|||||  
Db 130 GlnAsp 131  
RESULT 12  
AAG08682  
ID AAG08682 standard: Protein: 123 AA.  
XX AAG08682;  
AC AAG08682;  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6317.  
XX  
KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PT 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0128845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139492.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142390.  
PR 06-JUL-1999; 99US-0142055.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144086.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144884.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 28-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.

```

PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154039.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159283.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 22-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

```

```

Alignment Scores:
Pred. No.: 0.000143
Score: 117.50
Percent Similarity: 47.97%
Best Local Similarity: 30.08%
Query Match: 7.89%
DB: 21
Gaps: 5

US-09-763-902b-10 (1-830) x AAG08682 (1-123)

QY 158 GCTGAGGAGTTGTGCTACTGCTACACACCATGATAGCGGCGTTGCTGCC 217
      ..... ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 7 SerLysAlaPheValGluHisIleTyrSerThrPheAspThrAsnArgValGlyLeuAla 26
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 218 CGCGTGTACATGGCCACACCCCGTGTGGAATGCAATGCTTTCAGCAAGA 277
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 27 GlyLeuTyrGlnGluHisIleSerMetLeuThrPheGlnGlyGlnLysIleGlnGlyValGln 46
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 278 TCCTTGAGTGAAGTTTGAATGTTGCTT-----TCCAGCGAGTTCCAAATCAGCTG 331
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 47 SerIleValAlaLysLeuThrSerLeuProPheGlnGlnGlnCysIleHisIleSerThr 66
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 332 GTAGACTGCCAGCCTTGTATGATGAGCACACCAACCAACCGCTGCTTGTTC 391
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 67 ValAspCysGlnPro-----SerGlyProAlaSerGlyMetLeuValPhe 81
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 392 ATCTGTGATGATGAGTTGAGTTGAGGAGCAACCAACCGG---GACTTCAACCAAGCTTC 448
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 82 ValSerGlyAsnLeuGlnLeuAlaGlyGlnGlnHisAlaLeuLysPheSerGlnMetPhe 101
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 449 ATCTGTG-----ACCCGCCAGCGCTCACCACCAACCAACCAAGTGTGAGATCGCAAGTGC 502
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 102 HisLeuMetProThrProGlnGlnGlnSer-----PheTyrValPheAsnAsp 116
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 503 TGCTTCCGC 511
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 117 IlePheArg 119
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 13
AAG08681
ID AAG08681 standard; Protein; 126 AA.
XX AC AAG08681;
XX 17-OCT-2000 (first entry)
DE Arabidopsis thaliana protein fragment SEQ ID NO: 6316.
XX
XX Protein identification: signal transduction pathway; metabolic pathway;
KM hybridisation assay; genetic mapping; gene expression control; promoter;
KM termination sequence.
XX
XX Arabidopsis thaliana.
OS
XX EP1033405-A2.
PN
XX 06-SEP-2000.
PD
XX 25-FEB-2000; 2000EP-0301439.
PE
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125288.
PR 25-MAR-1999; 99US-0126284.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.

```

PR 23-APR-1999;	99US-0130891.
PR 28-APR-1999;	99US-0131449.
PR 30-APR-1999;	99US-0132048.
PR 30-APR-1999;	99US-0132407.
PR 04-MAY-1999;	99US-0132484.
PR 05-MAY-1999;	99US-0132485.
PR 06-MAY-1999;	99US-0132486.
PR 07-MAY-1999;	99US-0132863.
PR 11-MAY-1999;	99US-0134256.
PR 14-MAY-1999;	99US-0134218.
PR 14-MAY-1999;	99US-0134219.
PR 14-MAY-1999;	99US-0134221.
PR 18-MAY-1999;	99US-0134370.
PR 19-MAY-1999;	99US-0134768.
PR 20-MAY-1999;	99US-0135124.
PR 21-MAY-1999;	99US-0135353.
PR 24-MAY-1999;	99US-0135629.
PR 25-MAY-1999;	99US-0136021.
PR 27-MAY-1999;	99US-0136392.
PR 28-MAY-1999;	99US-0136782.
PR 01-JUN-1999;	99US-0137222.
PR 03-JUN-1999;	99US-0137528.
PR 04-JUN-1999;	99US-0137502.
PR 07-JUN-1999;	99US-0137724.
PR 08-JUN-1999;	99US-0138094.
PR 10-JUN-1999;	99US-0138540.
PR 10-JUN-1999;	99US-0138847.
PR 14-JUN-1999;	99US-0139119.
PR 16-JUN-1999;	99US-0139452.
PR 16-JUN-1999;	99US-0139453.
PR 17-JUN-1999;	99US-0139492.
PR 18-JUN-1999;	99US-0139454.
PR 18-JUN-1999;	99US-0139455.
PR 18-JUN-1999;	99US-0139456.
PR 18-JUN-1999;	99US-0139457.
PR 18-JUN-1999;	99US-0139458.
PR 18-JUN-1999;	99US-0139459.
PR 18-JUN-1999;	99US-0139460.
PR 18-JUN-1999;	99US-0139461.
PR 18-JUN-1999;	99US-0139462.
PR 18-JUN-1999;	99US-0139463.
PR 18-JUN-1999;	99US-0139750.
PR 18-JUN-1999;	99US-0139763.
PR 21-JUN-1999;	99US-0139817.
PR 22-JUN-1999;	99US-0139899.
PR 23-JUN-1999;	99US-0140353.
PR 23-JUN-1999;	99US-0140354.
PR 24-JUN-1999;	99US-0140695.
PR 28-JUN-1999;	99US-0140823.
PR 29-JUN-1999;	99US-0140991.
PR 30-JUN-1999;	99US-0141287.
PR 01-JUL-1999;	99US-0141842.
PR 01-JUL-1999;	99US-0142154.
PR 02-JUL-1999;	99US-0142055.
PR 06-JUL-1999;	99US-0142390.
PR 08-JUL-1999;	99US-0142803.
PR 09-JUL-1999;	99US-0142920.
PR 12-JUL-1999;	99US-0142977.
PR 13-JUL-1999;	99US-0143542.
PR 14-JUL-1999;	99US-0143624.
PR 15-JUL-1999;	99US-0144005.
PR 16-JUL-1999;	99US-0144085.
PR 16-JUL-1999;	99US-0144086.
PR 19-JUL-1999;	99US-0144325.
PR 19-JUL-1999;	99US-0144331.
PR 19-JUL-1999;	99US-0144332.
PR 19-JUL-1999;	99US-0144333.
PR 19-JUL-1999;	99US-0144334.
PR 19-JUL-1999;	99US-0144335.
PR 20-JUL-1999;	99US-0144352.
PR 20-JUL-1999;	99US-0144632.
PR 20-JUL-1999;	99US-0144884.
PR 21-JUL-1999;	99US-0144814.
PR 21-JUL-1999;	99US-0145086.
PR 21-JUL-1999;	99US-0145088.
PR 22-JUL-1999;	99US-0145085.
PR 22-JUL-1999;	99US-0145087.
PR 22-JUL-1999;	99US-0145089.
PR 22-JUL-1999;	99US-0145192.
PR 23-JUL-1999;	99US-0145145.
PR 23-JUL-1999;	99US-0145218.
PR 26-JUL-1999;	99US-0145224.
PR 27-JUL-1999;	99US-0145276.
PR 27-JUL-1999;	99US-0145913.
PR 27-JUL-1999;	99US-0145918.
PR 27-JUL-1999;	99US-0145919.
PR 28-JUL-1999;	99US-0145951.
PR 02-AUG-1999;	99US-0146386.
PR 02-AUG-1999;	99US-0146388.
PR 02-AUG-1999;	99US-0146389.
PR 03-AUG-1999;	99US-0147038.
PR 04-AUG-1999;	99US-0147204.
PR 04-AUG-1999;	99US-0147302.
PR 05-AUG-1999;	99US-0147192.
PR 05-AUG-1999;	99US-0147260.
PR 06-AUG-1999;	99US-0147303.
PR 06-AUG-1999;	99US-0147416.
PR 09-AUG-1999;	99US-0147493.
PR 09-AUG-1999;	99US-0147933.
PR 10-AUG-1999;	99US-0148171.
PR 11-AUG-1999;	99US-0148319.
PR 12-AUG-1999;	99US-0148341.
PR 13-AUG-1999;	99US-0148565.
PR 13-AUG-1999;	99US-0148684.
PR 16-AUG-1999;	99US-0149368.
PR 17-AUG-1999;	99US-0149375.
PR 18-AUG-1999;	99US-0149426.
PR 20-AUG-1999;	99US-0149722.
PR 20-AUG-1999;	99US-0149723.
PR 20-AUG-1999;	99US-0149929.
PR 23-AUG-1999;	99US-0149902.
PR 23-AUG-1999;	99US-0149930.
PR 25-AUG-1999;	99US-0150566.
PR 26-AUG-1999;	99US-0150884.
PR 27-AUG-1999;	99US-0151065.
PR 27-AUG-1999;	99US-0151066.
PR 27-AUG-1999;	99US-0151080.
PR 30-AUG-1999;	99US-0151303.
PR 31-AUG-1999;	99US-0151438.
PR 01-SEP-1999;	99US-0151930.
PR 07-SEP-1999;	99US-0152363.
PR 10-SEP-1999;	99US-0153070.
PR 13-SEP-1999;	99US-0153758.
PR 15-SEP-1999;	99US-0154018.
PR 16-SEP-1999;	99US-0154039.
PR 20-SEP-1999;	99US-0154779.
PR 22-SEP-1999;	99US-0155139.
PR 23-SEP-1999;	99US-0155486.
PR 24-SEP-1999;	99US-0155659.
PR 28-SEP-1999;	99US-0156458.
PR 29-SEP-1999;	99US-0156566.
PR 04-OCT-1999;	99US-0157117.
PR 05-OCT-1999;	99US-0157753.
PR 06-OCT-1999;	99US-0157865.
PR 07-OCT-1999;	99US-0158029.
PR 08-OCT-1999;	99US-0158232.
PR 12-OCT-1999;	99US-0158369.
PR 13-OCT-1999;	99US-0158293.
PR 13-OCT-1999;	99US-0158294.
PR 13-OCT-1999;	99US-0158295.
PR 14-OCT-1999;	99US-0158329.
PR 14-OCT-1999;	99US-0158330.
PR 14-OCT-1999;	99US-0158331.
PR 14-OCT-1999;	99US-0159637.

PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 28-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.: 0.000144 Length: 126  
Score: 117.50 Matches: 37  
Percent Similarity: 47.97% Conservative: 22  
Best Local Similarity: 30.08% Mismatches: 49  
Query Match: 7.89% Indels: 15  
DB: 21 Gaps: 5

US-09-763-902b-10 (1-830) x AAG08681 (1-126)

OY 158 GCGAGAGATGTGTCATGTCTACTACACCATGATGATAGCGGCGCTTGTCTCC 217  
DB 10 SerLysAlaPheValAluHisTyrTyrSerThrPheAsnArgValGlyLeuAla 29  
OY 218 CGCGCTTACATGGCGACACACACCTGTCTGATGCAATGCAATCTTTCAGGACAAGA 277  
DB 30 GlyLeuTyrGlnGluAlaSerMetLeuThrPheGlnGlyGlnArgLeuGln 49  
OY 278 TCCTTACGAGAGTTTGTGAATGTTGCTT-----TCCAGCGATGCCAATACACGCG 331  
DB 50 SerLeuValAlaLysLeuThrSerLeuProPheGlnGlnCysLysHisIleSerThr 69  
OY 332 GATCTGCGACCGCTTGTATGATGAGCACACACGACGACGCGCTTGTCTGTC 391  
DB 70 ValAspCysGlnPro-----SerGlyProAlaSerGlyMetLeuValPhe 84  
OY 392 ATCTGTGATCACTGAAGTTTGAAGGGAACAACAACGCG--GACTTCAACGAGACTTC 448  
DB 85 ValSerGlyAsnLeuGlnLeuAlaGlyGlnGlnHisAlaLeuLysHisSerGlnMetPhe 104  
OY 449 ATCTGTG-----ACCGCCCAAGCGCTTCAACGACACAGCAAGTGTGAGAGATGCAAGTGC 502  
DB 105 HisLeuMetProThrProGlnGlnGlySer-----PheTyrValPheAsnAsp 119  
OY 503 TGCTTCCGC 511  
DB 120 IlePheArg 122

## RESULT 14

AAG05147 standard; Protein; 131 AA.

AAG05147:

DT 17-OCT-2000 (first entry)

DE Arabidopsis thaliana protein fragment SEQ ID NO: 1443.

KW Protein identification: signal transduction pathway; metabolic pathway;  
hybridisation assay; genetic mapping; gene expression control; promoter;  
termination sequence.

XX XX Arabidopsis thaliana.  
OS  
XX  
XX EPI033405-A2.  
PN  
XX  
XX 06-SEP-2000.  
PD  
XX  
XX 25-FEB-2000; 2000EP-0301439.  
PE  
XX  
XX 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 23-APR-1999; 99US-0130891.  
PR 28-APR-1999; 99US-0131449.  
PR 30-APR-1999; 99US-0132048.  
PR 30-APR-1999; 99US-0132407.  
PR 04-MAY-1999; 99US-0132484.  
PR 05-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 06-MAY-1999; 99US-0132487.  
PR 07-MAY-1999; 99US-0132863.  
PR 11-MAY-1999; 99US-0134256.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.  
PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.

PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 01-JUL-1999; 99US-0142154.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143542.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 20-JUL-1999; 99US-0144684.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145192.  
PR 23-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.  
PR 11-AUG-1999; 99US-0148319.  
PR 12-AUG-1999; 99US-0148341.  
PR 13-AUG-1999; 99US-0148565.  
PR 13-AUG-1999; 99US-0148684.  
PR 16-AUG-1999; 99US-0149368.  
PR 17-AUG-1999; 99US-0149175.  
PR 18-AUG-1999; 99US-0149426.  
PR 20-AUG-1999; 99US-0149722.  
PR 20-AUG-1999; 99US-0149723.  
PR 20-AUG-1999; 99US-0149929.  
PR 23-AUG-1999; 99US-0149902.  
PR 23-AUG-1999; 99US-0149930.  
PR 25-AUG-1999; 99US-0150566.  
PR 26-AUG-1999; 99US-0150884.  
PR 27-AUG-1999; 99US-0151065.  
PR 27-AUG-1999; 99US-0151066.  
PR 27-AUG-1999; 99US-0151080.  
PR 30-AUG-1999; 99US-0151303.  
PR 31-AUG-1999; 99US-0151438.  
PR 01-SEP-1999; 99US-0151930.  
PR 07-SEP-1999; 99US-0152263.  
PR 10-SEP-1999; 99US-0153070.

PR 13-SEP-1999; 99US-0153758.  
PR 15-SEP-1999; 99US-0154018.  
PR 16-SEP-1999; 99US-0154039.  
PR 20-SEP-1999; 99US-0154779.  
PR 22-SEP-1999; 99US-0155139.  
PR 23-SEP-1999; 99US-0155486.  
PR 24-SEP-1999; 99US-0155659.  
PR 28-SEP-1999; 99US-0156458.  
PR 29-SEP-1999; 99US-0156596.  
PR 04-OCT-1999; 99US-0157117.  
PR 05-OCT-1999; 99US-0157753.  
PR 06-OCT-1999; 99US-0157865.  
PR 07-OCT-1999; 99US-0158029.  
PR 08-OCT-1999; 99US-0158232.  
PR 12-OCT-1999; 99US-0158369.  
PR 13-OCT-1999; 99US-0159293.  
PR 13-OCT-1999; 99US-0159294.  
PR 13-OCT-1999; 99US-0159295.  
PR 14-OCT-1999; 99US-0159329.  
PR 14-OCT-1999; 99US-0159330.  
PR 14-OCT-1999; 99US-0159331.  
PR 14-OCT-1999; 99US-0159637.  
PR 14-OCT-1999; 99US-0159638.  
PR 18-OCT-1999; 99US-0159584.  
PR 21-OCT-1999; 99US-0160741.  
PR 21-OCT-1999; 99US-0160767.  
PR 21-OCT-1999; 99US-0160768.  
PR 21-OCT-1999; 99US-0160770.  
PR 21-OCT-1999; 99US-0160814.  
PR 21-OCT-1999; 99US-0160815.  
PR 22-OCT-1999; 99US-0160980.  
PR 22-OCT-1999; 99US-0160981.  
PR 22-OCT-1999; 99US-0160989.  
PR 25-OCT-1999; 99US-0161404.  
PR 25-OCT-1999; 99US-0161405.  
PR 25-OCT-1999; 99US-0161406.  
PR 26-OCT-1999; 99US-0161359.  
PR 26-OCT-1999; 99US-0161360.  
PR 26-OCT-1999; 99US-0161361.  
PR 26-OCT-1999; 99US-0161920.  
PR 28-OCT-1999; 99US-0161992.  
PR 28-OCT-1999; 99US-0161993.  
PR 29-OCT-1999; 99US-0162142.

## Alignment Scores:

Pred. No.: 0.000147  
Score: 117.50  
Percent Similarity: 48.008  
Best Local Similarity: 29.608  
Query Match: 7.898  
DB: 21

Length: 131  
Matches: 37  
Conservative: 23  
Mismatch: 50  
Indels: 15  
Gaps: 5

US-09-763-902B-10 (1-830) x AAG05147 (1-131)

QY 158 GCGTACGAGAGTTGTCATGTCATGACACCATGATGATGAGGCGGCTTTCGTGTC 217  
DB 7 SerLyAlaPheValIuhIstYrSerThrPheSphrThsmlrYalGlyLeuAla 26  
QY 218 CGCCTGTACATGGGACACACCCCTGCTGGAATGCAATGCTTTCAGACAGAA 277  
DB 27 GlyLeuTyrgIngluAlaSerMetLeuThrPheGluGlyInlYshIleGIngluAla 46  
QY 278 TCCTTAGAGTGGTTTTCGAATGTCCTT-----TCCAGGAGTTCCAAATCAGCGT 331  
DB 47 SerIleValAlaLysLeuThrSerLeuProPheGInlGInlYshIshIstIleSerThr 66  
QY 332 GTACACTGCACGCTTTCATGATGAAGCAGACCAAGCAGACGAGGCTTGTGTC 391  
DB 67 ValAspCysGlnPro-----SerGlyProAlaSerGlyMetLeuValPro 81  
QY 392 ATCTGTGATCAGTACGATGTTTTCAGGAGACCAACAGCG---GACTTCAACAGAACTTC 448  
DB 82 ValSerClyAsnLeuGlnLeuAlaGlyGluYshIshIleValLeuYshSerGlnMetPro 101



OY 449 ATCTG-----ACGCGCCAGGCTCACCACACACAGTGTGAGATCCAGTGAC 502  
||| ||| ||| |||  
Db 102 HsLauMePrOthPrOclnlySer-----PhetYrYalpheasnasp 116  
OY 503 TCGTCCGCTCCAG 517  
||| |||  
Db 117 IlePheSerTrpArg 121  
RESULT 15  
AAG05146  
ID AAG05146 standard: Protein: 134 AA.  
XX AAG05146;  
AC  
XX  
DT 17-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana protein fragment SEQ ID NO: 1442.  
XX  
XX Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX Arabidopsis thaliana.  
OS  
XX  
PN EP103405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-0301439.  
XX  
PR 25-FEB-1999; 99US-0121825.  
PR 05-MAR-1999; 99US-0123180.  
PR 09-MAR-1999; 99US-0123548.  
PR 23-MAR-1999; 99US-0125788.  
PR 25-MAR-1999; 99US-0126264.  
PR 29-MAR-1999; 99US-0126785.  
PR 01-APR-1999; 99US-0127462.  
PR 06-APR-1999; 99US-0128234.  
PR 08-APR-1999; 99US-0128714.  
PR 16-APR-1999; 99US-0129845.  
PR 19-APR-1999; 99US-0130077.  
PR 21-APR-1999; 99US-0130449.  
PR 23-APR-1999; 99US-0130510.  
PR 28-APR-1999; 99US-0130891.  
PR 30-APR-1999; 99US-0131449.  
PR 04-MAY-1999; 99US-0132407.  
PR 05-MAY-1999; 99US-0132484.  
PR 06-MAY-1999; 99US-0132485.  
PR 06-MAY-1999; 99US-0132486.  
PR 07-MAY-1999; 99US-0132487.  
PR 11-MAY-1999; 99US-0132486.  
PR 14-MAY-1999; 99US-0134218.  
PR 14-MAY-1999; 99US-0134219.  
PR 14-MAY-1999; 99US-0134221.  
PR 14-MAY-1999; 99US-0134370.  
PR 18-MAY-1999; 99US-0134768.  
PR 19-MAY-1999; 99US-0134941.  
PR 20-MAY-1999; 99US-0135124.  
PR 21-MAY-1999; 99US-0135353.  
PR 24-MAY-1999; 99US-0135629.  
PR 25-MAY-1999; 99US-0136021.  
PR 27-MAY-1999; 99US-0136392.  
PR 28-MAY-1999; 99US-0136782.  
PR 01-JUN-1999; 99US-0137222.  
PR 03-JUN-1999; 99US-0137528.  
PR 04-JUN-1999; 99US-0137502.  
PR 07-JUN-1999; 99US-0137724.  
PR 08-JUN-1999; 99US-0138094.  
PR 10-JUN-1999; 99US-0138540.  
PR 10-JUN-1999; 99US-0138847.

PR 14-JUN-1999; 99US-0139119.  
PR 16-JUN-1999; 99US-0139452.  
PR 16-JUN-1999; 99US-0139453.  
PR 17-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139454.  
PR 18-JUN-1999; 99US-0139455.  
PR 18-JUN-1999; 99US-0139456.  
PR 18-JUN-1999; 99US-0139457.  
PR 18-JUN-1999; 99US-0139458.  
PR 18-JUN-1999; 99US-0139459.  
PR 18-JUN-1999; 99US-0139460.  
PR 18-JUN-1999; 99US-0139461.  
PR 18-JUN-1999; 99US-0139462.  
PR 18-JUN-1999; 99US-0139463.  
PR 18-JUN-1999; 99US-0139750.  
PR 18-JUN-1999; 99US-0139763.  
PR 21-JUN-1999; 99US-0139817.  
PR 22-JUN-1999; 99US-0139899.  
PR 23-JUN-1999; 99US-0140353.  
PR 23-JUN-1999; 99US-0140354.  
PR 24-JUN-1999; 99US-0140695.  
PR 28-JUN-1999; 99US-0140823.  
PR 29-JUN-1999; 99US-0140991.  
PR 30-JUN-1999; 99US-0141287.  
PR 01-JUL-1999; 99US-0141842.  
PR 02-JUL-1999; 99US-0142055.  
PR 06-JUL-1999; 99US-0142390.  
PR 08-JUL-1999; 99US-0142803.  
PR 09-JUL-1999; 99US-0142920.  
PR 12-JUL-1999; 99US-0142977.  
PR 13-JUL-1999; 99US-0143442.  
PR 14-JUL-1999; 99US-0143624.  
PR 15-JUL-1999; 99US-0144005.  
PR 16-JUL-1999; 99US-0144085.  
PR 16-JUL-1999; 99US-0144086.  
PR 19-JUL-1999; 99US-0144325.  
PR 19-JUL-1999; 99US-0144331.  
PR 19-JUL-1999; 99US-0144332.  
PR 19-JUL-1999; 99US-0144333.  
PR 19-JUL-1999; 99US-0144334.  
PR 19-JUL-1999; 99US-0144335.  
PR 20-JUL-1999; 99US-0144352.  
PR 20-JUL-1999; 99US-0144632.  
PR 21-JUL-1999; 99US-0144814.  
PR 21-JUL-1999; 99US-0145086.  
PR 21-JUL-1999; 99US-0145088.  
PR 22-JUL-1999; 99US-0145085.  
PR 22-JUL-1999; 99US-0145087.  
PR 22-JUL-1999; 99US-0145089.  
PR 22-JUL-1999; 99US-0145145.  
PR 23-JUL-1999; 99US-0145218.  
PR 23-JUL-1999; 99US-0145224.  
PR 26-JUL-1999; 99US-0145276.  
PR 27-JUL-1999; 99US-0145913.  
PR 27-JUL-1999; 99US-0145918.  
PR 27-JUL-1999; 99US-0145919.  
PR 28-JUL-1999; 99US-0145951.  
PR 02-AUG-1999; 99US-0146386.  
PR 02-AUG-1999; 99US-0146388.  
PR 02-AUG-1999; 99US-0146389.  
PR 03-AUG-1999; 99US-0147038.  
PR 04-AUG-1999; 99US-0147204.  
PR 04-AUG-1999; 99US-0147302.  
PR 05-AUG-1999; 99US-0147192.  
PR 05-AUG-1999; 99US-0147260.  
PR 06-AUG-1999; 99US-0147303.  
PR 06-AUG-1999; 99US-0147416.  
PR 09-AUG-1999; 99US-0147493.  
PR 09-AUG-1999; 99US-0147935.  
PR 10-AUG-1999; 99US-0148171.

PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0148341.  
 PR 13-AUG-1999; 99US-0148365.  
 PR 13-AUG-1999; 99US-0148684.  
 PR 16-AUG-1999; 99US-0149368.  
 PR 17-AUG-1999; 99US-0149175.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 20-AUG-1999; 99US-0149929.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-0154039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155486.  
 PR 24-SEP-1999; 99US-0155659.  
 PR 28-SEP-1999; 99US-0156458.  
 PR 29-SEP-1999; 99US-0156596.  
 PR 04-OCT-1999; 99US-0157117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0158029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 13-OCT-1999; 99US-0159285.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 14-OCT-1999; 99US-0159637.  
 PR 14-OCT-1999; 99US-0159638.  
 PR 18-OCT-1999; 99US-0159584.  
 PR 21-OCT-1999; 99US-0160741.  
 PR 21-OCT-1999; 99US-0160767.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160814.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160980.  
 PR 22-OCT-1999; 99US-0160981.  
 PR 22-OCT-1999; 99US-0160989.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.  
 PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161992.  
 PR 28-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Alignment Scores:

Pred. No.: 0.000149  
 Score: 117.50  
 Percent Similarity: 48.008  
 Best Local Similarity: 29.608  
 Query Match: 7.898  
 DB: 21

Length: 134  
 Matches: 37  
 Conservative: 23  
 Mismatches: 50  
 Indels: 15  
 Gaps: 5

US-09-763-902b-10 (1-830) x AAC05146 (1-134)  
 QY 158 GCTGAGAGTTTGTCAATGCTACTACACACCATGATAGCGCGCTTTCCTC 217  
 Db 10 SerLysAlaPheValIGluHISrYrYrSerThPheAspHrsmAtgValIGlyLeuAla 29  
 QY 218 CGCCTGTACATGGGACACACCGCCACCTGCTGGAATGCAATGCTTTCCAGACAGAA 277  
 Db 30 GlyLeuYrGlnGluAlaSerMetLeuThrPheGlnGlyGlnGlyGlnGlyValGln 49  
 QY 278 TCCTTGAGTGAAGTTTGAATGTCCT-----TCACGCGAGTTCACAAATCAGCGTG 331  
 Db 50 SerIleValAlaLysLeuThrSerLeuProPheGlnGlnGlySerHisHisIleSerThr 69  
 QY 332 GTAGACTGCCAGCCTGTTTCATGATGAAGCCACACACACAGCCAGCAGGCTTGTGTC 391  
 Db 70 ValAspCysGlnPro-----SerIlyProAlaSerGlyMetLeuValPhe 84  
 QY 392 ATCTGTGATCAGTGAAGTTTGAAGGGAACAAACACAGG---GACTTCACACAGACTTC 448  
 Db 85 ValSerGlyAsnLeuGlnLeuAlaGlyGlnHisHisAlaLeuLysPheSerGlnMetPhe 104  
 QY 449 ATCTGTG-----ACCGCCAGCGCTTCACCCAGCACACAGTGTGAAGATCGCAAGTGAC 502  
 Db 105 HisLeuMetProThrProGlnGlySer-----PheTyrValPheAsnAsp 119  
 QY 503 TGCCTCCGCTTCAG 517  
 Db 120 IlePheSerTrpArg 124

Search completed: June 24, 2003, 19:31:03  
 Job time : 43.6938 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 24, 2003, 19:27:58 ; Search time 10.711 Seconds  
(without alignments)  
4559.976 Million cell updates/sec

Title: US-09-763-902B-10  
Perfect score: 1489  
Sequence: 1 tgaacttgacattcactg.....ataatataataacacatg 830

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 262574 seqs, 29422922 residues  
Total number of hits satisfying chosen parameters: 525148

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame-n2p.model -DEV=xlp  
-O/cgn2.1/USPTO.spool/US09763902/rnat\_24062003.130342.12495/app-query.fasta.1.2254  
-DB=Issued Patents.AA -OFMT=fasta -SUFFIX=n2p.ra1 -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=haman40.cdf  
-LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFM=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09763902.cgn1.1.42.gnat\_24062003.130342.12495 -NCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FEAPEXT=7 -YGAPOP=10 -YGAEXT=0.5 -DELOP=6 -DELEXT=7

Database: Issued Patents.AA.\*

- 1: /cgn2.6/prodata/1/aa/5A.COMB.pep.\*
- 2: /cgn2.6/prodata/1/aa/5B.COMB.pep.\*
- 3: /cgn2.6/prodata/1/aa/5A.COMB.pep.\*
- 4: /cgn2.6/prodata/1/aa/5B.COMB.pep.\*
- 5: /cgn2.6/prodata/1/aa/PCTUS.COMB.pep.\*
- 6: /cgn2.6/prodata/1/aa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	90	6.0	466	2	US-08-836-791-9
2	88.5	5.9	671	4	US-08-858-207A-301
3	81.5	5.5	498	1	US-08-660-963-12
4	80	5.4	1276	1	US-08-222-616-24
5	80	5.4	1276	4	US-08-446-648-24
6	80	5.4	1276	5	PCT-US95-04228-24
7	78.5	5.3	492	4	US-09-413-814-72
8	78.5	5.3	2972	4	US-09-579-181-2
9	78.5	5.3	3118	4	US-09-579-181-1
10	78	5.2	480	4	US-08-570-227A-2
11	78	5.2	480	4	US-09-077-091-2
12	78	5.2	1098	1	US-08-290-301-82

13	78	5.2	1098	4	US-09-013-598-82
14	77	5.2	838	2	US-08-216-260-4
15	76.5	5.1	766	2	US-08-846-762-5
16	76	5.1	301	4	US-09-231-227-4
17	75	5.0	521	2	US-08-682-847-4
18	75	5.0	752	4	US-09-402-929-6
19	74.5	5.0	217	4	US-09-602-543-5
20	74.5	5.0	235	4	US-09-602-543-4
21	74.5	5.0	253	4	US-09-154-750A-83
22	74.5	5.0	1081	4	US-09-369-364A-17
23	74.5	5.0	1251	5	PCT-US95-02251-3
24	74.5	5.0	1252	1	US-08-199-780-3
25	74.5	5.0	1252	2	US-08-316-650-3
26	73.5	4.9	434	4	US-09-540-245A-19
27	73.5	4.9	1082	4	US-09-336-643A-20
28	73.5	4.9	3066	4	US-08-952-127-12
29	73	4.9	310	2	US-08-943-600A-3
30	73	4.9	502	1	US-08-278-635B-7
31	73	4.9	502	3	US-08-464-258B-7
32	73	4.9	502	3	US-08-471-961-7
33	73	4.9	688	3	US-08-941-445A-11
34	72.5	4.9	587	3	US-08-899-578-2
35	72.5	4.9	801	1	US-08-725-012-2
36	72.5	4.9	899	4	US-09-422-936-71
37	72.5	4.9	902	4	US-09-644-600-10
38	72.5	4.9	1037	4	US-09-428-711A-21
39	72.5	4.9	1083	4	US-09-600-776-2
40	72.5	4.9	1083	4	US-09-343-494-1
41	72	4.8	163	2	US-08-727-688-25
42	72	4.8	297	3	US-08-651-136C-4
43	72	4.8	297	4	US-09-229-911A-4
44	72	4.8	519	4	US-09-172-841-55
45	72	4.8	1754	1	US-07-745-206A-13

## ALIGNMENTS

RESULT 1  
US-08-836-791-9  
Sequence 9, Application US/08836791  
Patent No. 5886150  
GENERAL INFORMATION:  
APPLICANT: Duchesne, Marc  
APPLICANT: Faucher, Didier  
APPLICANT: Parker, Fabienne  
APPLICANT: Schweighofer, Fabien  
APPLICANT: Tocque, Bruno  
TITLE OF INVENTION: Peptides Capable of Binding to the GAP  
TITLE OF INVENTION: Protein SH3 Domain, Nucleotide Sequences Coding Thereof  
NUMBER OF SEQUENCES: 10  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcola Road, Mailstop 3043  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19426  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION NUMBER: US/08/836,791  
FILING DATE:  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 94/13955  
FILING DATE: 22-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 95/05753  
FILING DATE: 16-MAY-1995



Score:	81.50	Matches:	53
Percent Similarity:	33.33%	Conservative:	16
Best Local Similarity:	25.60%	Mismatches:	52
Query Match:	5.47%	Indels:	86
DB:	2	Gaps:	10

US-09-763-902B-10 (1-830) x US-08-660-963-12 (1-498)

QY	159	CTGAGAGATTGTCAATGCTACTACACCAACCAGATGATACGGC-----	203
DB	40	LeuthrglnAlaAlaValAlaGlnUleuProProtrpAlaAlaGlyCysGlyAlaProAla	59
QY	204	---GGCGTTGCTGTGCCCCCTGT-----ACATGGGGA-----	233
DB	60	SerSerAlaCysTrpAlaProCysGlnSerSerTrpAlaTrnCysThrGlnSerValMet	79
QY	234	-----	CAG 236
DB	80	SerSerLeuSerGlnArgThrSerIshsValTyrIlysLeuLeuGlyCysProThr	99
QY	237	CCACCTCG-----TCGGAATGGCAATGCTGTTTCAGAGACAGAACTCTTGA	284
DB	100	ProProTrpAlaAlaProGlySerGlyThrGly-CysCys-----	112
QY	285	GtGAGATTTTGAATGTGGCTTCACAGGAGTCCAAATCAGCGTGAGACTCCAGC	344
DB	113	-----AlaGlyArg-TTrpGlnA	118
QY	345	CTGTTCATGATGAAGCCACACCAAGCAGACCGGCTCTTGTTGCATCTGTGATCAG	404
DB	118	IaLeuGlnSerIlyAlaSerPro-AlaArgLeuSerSerLeuThrSerIaTrpSerGln	137
QY	405	TGAAGTTTGAGGGGAGCAAAACAAGGGGACTTCACCAACT---TCATCTGACCGCCC	461
DB	138	---GlyLeuArgIlyThrAlaProLeuGlnAlaGlyArgSerProSerArgLeuIlePro	156
QY	462	AGGCGT-----CACCAAGCAACAGAGTGGGAAGA	491
DB	157	ArgProIaIaLeuCysProTrpSerCysLeuArgAlaGlnProThrSerProArgIlySer	176
QY	492	TCGCAGATGATGCTCCGCTTCACAGACTGGGCGCAGCT-----	530
DB	177	SerThrProTrpAlaThrAlaSerArgLeuGlnProSerTrpProSerSerItrpSer	196
QY	531	---AGTGGGGGGGAGAGAGGTCTTTGCTTCATTCAGCCCTAGCTCTGTAGAGAAATGC	587
DB	197	LeuSerGlyGlySerThrAlaProGlyThrThrSerThrProSer-----Cys	212
QY	588	AAACTCGACTCTCA 602	
DB	213	SerProProLeuSer 217	

RESULT 4  
US-08-222-616-24  
Sequence 24, Application US/08222616  
Patent No. 5635177

GENERAL INFORMATION:  
APPLICANT: Bennett, Brian D.  
APPLICANT: Goeddel, David  
APPLICANT: Lee, James M.  
APPLICANT: Matthews, William  
APPLICANT: Tsai, Siao Ping  
APPLICANT: Wood, William I.  
TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST  
TITLE OF INVENTION: ANTIBODIES  
NUMBER OF SEQUENCES: 42  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080

Page 4

US-09-763-902B-10 (1-830) x US-08-446-648-24 (1-1276)

QY	33	CCTTACTCTCCCTGCAGACCCCTGGTTCCTCCCAAGGACAGAGAAATACCT	---GGTGGAGCC	89
Db	985	ProGlnIlyrProAlaGlyIthrProHlaIsrProArGyspThrAlaSerProPheSerGlyAla		1004
QY	90	CTCTCTCCATAGAAACAGAGATGGCATCTGTGGATTTCAAGA	---CCATATGG	140
Db	1005	-----GluTPrGlyLeuThrGluAlaProSerProValArgTyr		1018
QY	141	ATCAGGCCCTGCAGAGCTGCTGAGAGATTTTGCAATGTCTACTACACCCAGCATGATTAAGC	200	
Db	1019	IlleAlaLeuAlaArgGlyAlaArgSer	-----TyrGlnPhe	1030
QY	201	GGCGGCGCTTGGCTGTCCCGCCCTGTACATGGGACACAGCCACCCCTGGTGTGGANTGCAATG	260	
Db	1031	Gly	-----GluThrGlyPheGlyGlySerAlaIle	1040
QY	261	CTGTTCAGGCAAGAAATCCTTAGTAGTGTTTTGAATGTTCCTTCAGGACAGATTC	320	
Db	1041	IlleGlyGlyGluAsnHisPro	-----ProAlaThrSer	1051
QY	321	AAATCAGCGGTGTAGACTGCCAGCCCTGTTTCATGATGAAGCCACACCAAGCCAGACCA	--377	
Db	1052	GlyAsnSer	-----ArgProArgValAlaGalaProPhe	1062
QY	378	-----CGGTCCCTGTTCATCTGTGGATCAGTACGTGAAGTTTAGG	416	
Db	1063	ProGlnAspTyrPvalProGluGluIlySerGluValProAlaIleSerGln	-----1078	
QY	417	GGAACAAACAAAGGGACTTCAACACAGACTTCATCTGACCGCCAGG	---CTTCAC	470
Db	1079	-----ProProGlnValProProSer	1085	
QY	471	CCAGAACACAGTGTGAAGA	---TCCGAAGTACGTGCTCCGCTTCAGAGCTGGGCCA	527
Db	1086	ProTyrPvalArgSerArgArgProIlyAspValIleuProCysGlnLeuGlnSerGlyIly	1105	
QY	528	GCTAGTGGGGGTGGCAGAGGTCTCTTCTTCATTCACGCCCTAGTCTGTAGAGAAATGC	587	
Db	1106	AlaValProGlyGlyLys	-----LysGlyLys	1114
QY	588	AAACCTGCAGCTCTCAAGATGTGAGAACACACAGTTCATTTCTGTGTGTGGAGACACT	647	
Db	1115	GlnGlyProValThrLysSerLeuGlyPheValValProIlnCysCysHisHisGln	1134	
QY	648	GGAGACTCCACGTGCCAGGTGAACCT	677	
Db	1135	ThrGlnSerPhePheSerLeuValAsnAlaProProAlaAlaAlaPheIleLeuLys	1155	
QY	678	-----TTTTTGTGCTCAAGTCTTAGAGAGTCCCTTCTCGTAGATATATACTGTT	726	
Db	1155	ValPheGluPheCysPheTyrPhePheSerProPheProPheCysPhePheValle	1174	
QY	727	-----TGTCAATGTTCTCTTTTCAAGAGTAAAGTATTTTCT	770	
Db	1174	uPhePheTyrArgProCysHisAsnPheValIleuGluGlyIthrCysPheThrMetalase	1194	
QY	771	ACTGGCCAG	780	
Db	1194	rPheAlaGln	1197	

RESULT 6

```

GENERAL INFORMATION:
APPLICANT: Genentech, Inc.
APPLICANT: Bennett, Brian D.
APPLICANT: Goeddel, David
APPLICANT: Lee, James M.
APPLICANT: Matthews, William
APPLICANT: Tsai, Siao Ping
APPLICANT: Wood, William I.
TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES
NUMBER OF SEQUENCES: 45
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/04228
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/222616
FILING DATE: 04-APR-1994
ATTORNEY/AGENT INFORMATION:
NAME: Wendy M. Lee
REGISTRATION NUMBER: 00,000
REFERENCE/DOCKET NUMBER: 821P3PCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-1994
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 24:
SEQUENCE CHARACTERISTICS:
LENGTH: 1276 amino acids
TYPE: amino acid
TOPOLOGY: linear
PCT-US95-04228-24

Alignment Scores:
Pred. No.: 1.43 Length: 1276
Score: 80.00 Matches: 59
Percent Similarity: 32.75% Conservative: 34
Best Local Similarity: 20.77% Mismatches: 86
Query Match: 5.37% Indels: 105
Gaps: 14
DB:

US-09-763-9028-10 (1-830) x PCT-US95-04228-24 (1-1276)
33 CCTTACTTCCTGACAGCCCTGTTCCCAAGCAGAGAAATACCT---GGTGAGCC 89
|||:::||||| ||||| ||| |||||
985 PROGINTYPRVALAGLYTHRPROHSPRODAGAPTHALASERPROHESERGIALA 100
Query Match: 5 Gaps: 14

1005 -----GLUTRPGLYLEUTHRGUALAProSerProValProArgTrp 1011
||| ||| ||| ||| |||
141 ATCAGGCGTCGACAGAGCTGTGAGAGATTGTCAATGTCTACTACACACCATGATGAAC 200
||| ||| ::||| ||| |||
1019 ILEALALEUALARGLYVALARGSER-----TRPGLNPro 1030
||| ||| ||| ||| |||
201 GGGCGCGTTTGTCTGCCGCGCTGTACATGAGGACACAGCCACTGCTGTGGAATGCGCATG 260
||| ||| ||| ||| |||
1031 GLY-----GLUTHRGlyPheGLYGLYSERLALILE 1040
||| ||| ||| ||| |||
261 CCGTTTCAGACAGAAATCCTTGAGAGATTTTTTGAATATGTTGCGTCCACAGCGATTCC 320

```

```

Db      1041  illegalgylusnhispro-----ProLathrser 1051
Oy      321  AATTCAGCGGTGAGACTGCCACCGCTGTCATGATGAGAACCCACACCCAGACCA--- 377
Db      1052  Glysenser-----ArgProArgValArgLalrophe 1066
Oy      378  -----CGGTCCTTGTTGATCTGTGGATCAGTAGAATTGGAG 416
Db      1063  ProGlnAspTrpValProGlnLulysGluValProAsnLlserGln----- 1078
Oy      417  GGAAACAACAGGGAGCTTCACACAGAACTTATCTGACCCGCCAG-----CCTCAC 470
Db      1079  -----ProGlnIvalProProser 1085
Oy      471  CCAGACACACAGCTGTGAGA---TCGCAGAGTACTGCTCCGCTCCAGGACTGGGCCA 527
Db      1086  ProTrpValArgSerArgItrProLysArgValLeuProCysGlnLeuGlnSerLylGly 1105
Oy      528  GCTAGTGGGGGTGGCAGAGGCTCTTTGCTTCATTCAAGCCCTAGCTGTAGAGAAATGC 587
Db      1106  AlalaValProGlnLylGlys-----LysGlyCys 1114
Oy      588  AATCTCGACTCTCAAGATGTGAGAACACAACTCATTTCTGTGTGGCGAGACACT 647
Db      1115  GlnGlyProValThrLysSerLeuLylheValValProThrCysCysGlnLlserGln 1134
Oy      648  GCAGACTCCACTGCGCAGAGTTGAACCT----- 677
Db      1135  ThrGlnSerPhePheSerLeuValAsnLalProProAlaIalAlaPheLlLeuLys 1154
Oy      678  -----TTTTTCTCTCAAGTTCTAGAGTCCCTTCTCGAATATATACTGTT-- 726
Db      1155  ValPheGluPhe-CysPheTrpSerPhePheSerProPheSerProPheSerPheValle 1174
Oy      727  -----TGTCACTAGTTCTCTTCAAGTAACTAACTTTTCTATTTTCT 770
Db      1174  uPhePheTrpArgProCysHisAsnPheValLeuGlnGlyThrCysPheThrMetAlase 1194
Oy      771  ACTGCCCGAG 780
Db      1194  rPheAlaGln 1197

RESULT 7
US-09-413-814-72
; Sequence 72, Application US/09413814
; Patent No. 6225064
GENERAL INFORMATION:
APPLICANT: Gesellschaft fuer Biotechnologische Forschung mbH
APPLICANT: Bristol-Myers Squibb, Co.
APPLICANT: Beyer, Stefan
APPLICANT: Bloecker, Helmut
APPLICANT: Brandt, Petra
APPLICANT: Cino, Paul M
APPLICANT: Dougherty, Brian A
APPLICANT: Goldberg, Steven L
APPLICANT: Hoffe, Gerhard
APPLICANT: Mueller, Joachim
APPLICANT: Reichenbach, Hans
TITLE OF INVENTION: DNA sequences for enzymatic synthesis of polyketide or
FILE REFERENCE: PCT/US 99/23535
CURRENT APPLICATION NUMBER: US/09/413,814
CURRENT FILING DATE: 1999-10-07
EARLIER APPLICATION NUMBER: DE 198 46 493.2
EARLIER FILING DATE: 1998-10-09
NUMBER OF SEQ ID NOS: 107
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 72
LENGTH: 492
TYPE: PRT
ORGANISM: Sorangium cellulosum
US-09-413-814-72

```

Alignment Scores:	
Pred. No.:	1,31
Score:	78.50
Percent Similarity:	38.94%
Best Local Similarity:	30.09%
Query Match:	5.27%
DB:	4
	Gaps: 5
	Length: 492
	Matches: 34
	Conservative: 10
	Mismatches: 52
	Indels: 17
	Gaps: 5

```

OY      309  CAGAGCAGTTCAAATCACCCTGGTGAAGACTGCCACCCTGTTTCATGATGATACCCACACCAA 366
Db      28  ProTthProAlrgProAlrgIlyTThpAlrgTThrThsSerPethrThSerAlrgAlrgProSerAla 47
OY      369  GCAGAGCCACAGG-----TCCTTGTGTGATCTGGATCACTGAATCTTTAAGGGGA 419
Db      48  SerAlrgProAlaAlaSerIySerIySerIySerIySerAlrgAlaProSerProAlrgIly 67
OY      420  ACAAAACAGGGAGCTTCAACACAGAAT-----TCATCTCCAGCCGCCAG 464
Db      68  AsnGlySerGIuCySerThrThrThrThleuProAlaGIuGIuAlaSerTThpProAlrg 87
OY      465  CCT-----CACCCAGACACACAGTGT-----GGAGATCGCACAGTGACTGCTCC 509
Db      88  ProCyAlrgMetProTThrSerIySerIySerIySerProSerAlrgSerAlrgAlaAlaLeuSer 107
OY      510  GCTTCAGAGACTGGGCCAGCTAGTGGG-----GGTGGCAGAGGCTCTTTGCT 557
Db      108  LeuHISLeuLeuProProSerSerIyAlrgAlrgProSerGIyAsnIleAlaLeuSerAla 127
OY      558  TCATTTCAGCCCTAGCTGTGTAGAGAAAGCAAACTGCA 596
Db      128  AlaLeuSerProProAlaGlyProAlrgAlaLeuProAlrg 140

```

```

RESULT 8
US-09-579-181-2
: Sequence 2, Application US/09579181
: Patent No. 6365372
:
GENERAL INFORMATION:
: APPLICANT: Chiviva, John
: APPLICANT: Yaciuk, Peter
: TITLE OF INVENTION: SRF2 Related CBP Activator Protein (SRCAP)
: FILE REFERENCE: 16153-4247
: CURRENT APPLICATION NUMBER: US/09/579,181
: CURRENT FILING DATE: 2000-05-25
: PRIOR APPLICATION NUMBER: 60/136,620
: PRIOR FILING DATE: 1999-05-27
: NUMBER OF SEQ ID NOS: 17
: SOFTWARE: Patentlin Ver. 2.0
:
SEQ ID NO 2
: LENGTH: 2972
: TYPE: PRT
: ORGANISM: Human
US-09-579-181-2

Alignment Scores:
Pred. No.:
Score: 3.38 Length: 2972
Percent Similarity: 78.50 Matches: 55
Best Local Similarity: 32.17% Conservative: 19
Query Match: 23.918 Mismatches: 88
DB: 5.27% Indels: 68
Gaps: 7

US-09-763-902B-10 (1-830) x US-09-579-181-2 (1-2972).
27 CTTCAACCTTACTCTCCCTGACAGCCCTGTTCCCAAGACAGAGCAATACCTGGATGA 86
||| ||| ||||| ||||| ||| ||| ||| |||
1221 leuAlaProValValProAlaIaIaProGlyProProSeIeuGln-----ProSerGly 1238
87 GCCCTCTTCCATTAGAACCAAGATGCATCTGTGATTCAAAGACCTATGTGATCAAG 146
|||
1239 Ala-----Ser 1240

```



147 CCTGCAGAGCTGCTGAGAGTTGTCATGCTACTACCA----- 188  
1241 ProserAlaSerAlaLeuThrLeuGlyLeuAlaThrAlaProSerLeuSerSerSerIn 1260  
189 -----CCATGATTAAGCGC 203  
1261 ThrProGlyHisProLeuLeuLeuAlaProThrSerSerHisValProGlyLeuAnsSer 1280  
204 GGCCTTTGCTGTCCTCCCTGTACATGGGCACAGCCACCTGCTGCAATGGCATGCTG 263  
1281 ThrVal---AlaProAlaCysSerProValLeuValProAlaSerAlaLeuAlaSerPro 1299  
264 TTTCAGAGC---AAGATTCCTGAGTGGATTTTGAATGTCCTCCAGCGAGTCC 320  
1300 PheProSerAlaProAnsProAlaProAlaGlnAlaSerLeuLeuAlaProAlaSer 1319  
321 AATCA----- 326  
1320 AlaSerGlnAlaLeuAlaThrProLeuAlaProMetaAlaAlaProGlnThrAlaLeu 1339  
327 -----GGGTGTAGACTGCCAGCCTGTTCATGATGAGCAACCAACCAAGCA 374  
1340 AlaProSerProAlaProAlaProLeuAlaProLeuProValLeuAlaProSerProGlyAla 1359  
375 CCAGGCTCCTGTGTCTCTGTGATCAGTGAAGTTGAGGGGAAACAAC----- 428  
1360 AlaProValLeuAlaSerSerGlnThrProValProValMetaAlaProSerSerThrPro 1379  
429 GGGACTTCACACCACTTCATCTGACGCCGCCAGCTCAACCAACACAGAGTGA 488  
1380 GlyThrSerLeuAlaSerAlaSerProValProAlaProThrProValLeuAlaProSer 1399  
489 AGATCGCAAGTACTGCTCCGCTTCAGAGACTGGGCCACTAGTGGGGTGGCAGAGT 548  
1400 SerThrGlnThr-MetLeuProAlaProValProSerProLeuProSerProAlaSerTh 1419  
549 CTCCTTGTCTCATTCAGCCCTAGCTCTG 576  
1419 rGlnThrLeuAlaLeuAlaProAlaLeu 1428

RESULT 9  
US-09-579-181-1  
Sequence 1, Application US/09579181  
Patent No. 6365372  
GENERAL INFORMATION:  
APPLICANT: Chiviva, John  
APPLICANT: Yaciuk, Peter  
TITLE OF INVENTION: SMF2 Related CBP Activator Protein (SRCAP)  
FILE REFERENCE: 16153-4247  
CURRENT APPLICATION NUMBER: US/09/579,181  
CURRENT FILING DATE: 2000-05-25  
PRIOR APPLICATION NUMBER: 60/136,620  
PRIOR FILING DATE: 1999-05-27  
NUMBER OF SEQ ID NOS: 17  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1  
LENGTH: 3118  
TYPE: PRT  
ORGANISM: Human  
US-09-579-181-1

Alignment Scores:  
Pred. No.: 3 47 Length: 3118  
Score: 78.50 Matches: 55  
Percent Similarity: 32.17% Conservative: 19  
Best Local Similarity: 23.91% Mismatches: 88  
Query Match: 5.27% Indels: 68  
DB: 4 Gaps: 7

US-09-763-902b-10 (1-830) x US-09-579-181-1 (1-3118)  
27. CTCACAGCTTACTTCCCTGACGCCCTGTGTTCCCAAGGACAGAAATACCTGTGTA 86

1367 LeuAlaProValAlaProAlaAlaProGlyProProSerLeuGln-----ProserGly 1384  
87 GCCCTCTCCATAGAACAGAGATGGCATGTGTGATTTCAACACTATGGATCAGG 146  
1385 Ala-----Ser 1386  
147 CCTGCAGAGCTGCTGAGAGTTTTCATGCTACTACCA----- 188  
1387 ProserAlaSerAlaLeuThrLeuGlyLeuAlaThrAlaProSerLeuSerSerSerIn 1406  
189 -----CCATGATTAAGCGC 203  
1407 ThrProGlyHisProLeuLeuLeuAlaProThrSerSerHisValProGlyLeuAnsSer 1426  
204 GGCCTTTGCTGTCCTCCCTGTACATGGGCACAGCCACCTGCTGCAATGGCATGCTG 263  
1427 ThrVal---AlaProAlaCysSerProValLeuValProAlaSerAlaLeuAlaSerPro 1445  
264 TTTCAGAGC---AAGATTCCTGAGTGGATTTTGAATGTCCTCCAGCGAGTCC 320  
1446 PheProSerAlaProAnsProAlaProAlaGlnAlaSerLeuLeuAlaProAlaSer 1465  
321 AATCA----- 326  
1466 AlaSerGlnAlaLeuAlaThrProLeuAlaProMetaAlaAlaProGlnThrAlaLeu 1485  
327 -----GGGTGTAGACTGCCAGCCTGTTCATGATGAGCAACCAACCAAGCA 374  
1486 AlaProSerProAlaProAlaProLeuAlaProLeuProValLeuAlaProSerProGlyAla 1505  
375 CCAGGCTCCTGTGTGATCAGTGAAGTTGAGGGGAAACAAC----- 428  
1506 AlaProValLeuAlaSerSerGlnThrProValProValMetaAlaProSerSerThrPro 1525  
429 GGGACTTCACACCACTTCATCTGACGCCGCCAGCTCAACCAACACAGAGTGA 488  
1526 GlyThrSerLeuAlaSerAlaSerProValProAlaProThrProValLeuAlaProSer 1545  
489 AGATCGCAAGTACTGCTCCGCTTCAGAGACTGGGCCACTAGTGGGGTGGCAGAGT 548  
1546 SerThrGlnThr-MetLeuProAlaProValProSerProLeuProSerProAlaSerTh 1565  
549 CTCCTTGTCTCATTCAGCCCTAGCTCTG 576  
1565 rGlnThrLeuAlaLeuAlaProAlaLeu 1574

RESULT 10  
US-08-570-227A-2  
Sequence 2, Application US/08570227A  
Patent No. 5981217  
GENERAL INFORMATION:  
APPLICANT: Subramaniam, M.  
APPLICANT: Spelsberg, T. C.  
TITLE OF INVENTION: DNA ENCODING TGF-BETA INDUCIBLE  
TITLE OF INVENTION: EARLY FACTOR-1 (TIEF-1), A GENE EXPRESSED  
BY OSTEOBLASTS  
NUMBER OF SEQUENCES: 11  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Schwegman, Lundberg, Woessner & Kluth, P.A.  
STREET: P.O. Box 2938  
CITY: Minneapolis  
STATE: MN  
COUNTRY: USA  
ZIP: 55402  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
COMPUTER: IBM Compatible  
OPERATING SYSTEM: DOS  
SOFTWARE: FastSeq Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/570,227A  
FILING DATE: 11-DEC-1995

CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER:  
 FILING DATE:  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Moessner, Warren D  
 REGISTRATION NUMBER: 30,440  
 REFERENCE/DOCKET NUMBER: 150,157US1  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 612-359-3260  
 TELEFAX: 612-359-3263  
 TELEX:  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 480 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-570-227A-2

Alignment Scores:  
 Pred. No.: 1,49  
 Score: 78.00  
 Percent Similarity: 33.57%  
 Best Local Similarity: 23.21%  
 Query Match: 5.24%  
 DB: 2  
 Gaps: 14

US-09-763-902B-10 (1-830) x US-08-570-227A-2 (1-480)

```

OY 11 CATTACGCGGCTCTCTTCAACCTTACTCC-----TCAGCCCGCTGGTCCCAAG 64
DB 118 Hispelyserleuserprhralalysprohlsilaalalapro--Pheylsglu 136
OY 65 GCAGAGAAATACCGCTGGAGCCCTCTTCATGAAACGACATGTCATGTGGAT 124
DB 137 Glugluysserprovalserlarpolylsleuprolysalaglnlathrservalile 156
OY 125 TTCAGACCTATGTGATCAGCGCTGCAGA-----GCTGCTAGAGAG 166
DB 157 ArgHsthrAlaAspAlaglnleucysasnHsglnthrcysPrometlyslalaaser 176
OY 167 TTTCGCAATGTCTACTACACCATGATAGCGGCGGCTTG-----CTGTCC 217
DB 177 IleleuasnTyrglnAsnAsnSerPheargTargTgthHlsleuAsnValgluAlaala 196
OY 218 CGCGTGTACGGGACAGCCACCTGCTGTGATGCAATGCT----- 262
DB 197 ArglyAsnIleProcyAlaAlaValSerProAsnArgserlyscysgluArgasnThr 216
OY 263 -----GTTTCAGACAGAAATCCTTGAGTGAAGTTTGGTAATGTGCTCCAGC 313
DB 217 ValAlaAspValaAspGlnlyslaserAlaAlaLeuTyrgAspPheSerValaProserSer 236
OY 314 GAGTTCCAAATCAGCGGTAGACTGCCAGCTGTTCATGATGAGACCAACCAAGCAG 373
DB 237 Glu-----ThValIlecyAsrgserGlnProAlaProValSerProIndn 252
OY 374 ACCAGGCTCTGT-----GTC 391
DB 253 LysSerValleuValSerProProAlaValSerAlaglyValProPromeProVal 272
OY 392 ATCTGTGATCAGTGAAGTTTGAGGGAACAACAAGGAGCTTCACCAAGCTTCATC 451
DB 273 IlecySglMetValProleuProAlaAsn-AsnProValValIndnThrValProse 292
OY 452 CTCAGCGCCCAAG----- 464
DB 292 fThProProserGlnProProAlaValcysProProValValPheMetglYthnGlnVa 312
OY 465 -----CTTCACCCAGCAACACAGTGTGGAAGTCCCA 496
DB 465 -----CTTCACCCAGCAACACAGTGTGGAAGTCCCA 496

```

312. IProlYsGlyAlaValMetPheValValProGlnProValValGlnSerSerLysProPr 332

OY 497 AGTACGCGT-----TCCGCTTCAGACAGCGGCGAGCTGAGGGGTGCAG 544

DB 332 ovalValSerProAsnGlythArgleuSerProIleAlaProAlaProly----- 349

OY 545 AGCTCTCTTGTTCATTACAGCCCTAGCTGTAGAAATGCAACCTGCAATC----- 599

DB 350 -----PheSerProSerAlaala---LysValIndnProGlnIleAspse 363

OY 600 -TCAGAGTGTAGAAACAATTCATTCTCTGT-----TCGAGACACT 647

DB 363 rSerArgIle-----ArgSerHlsIlecySerHisProGlyCysGlyLysThr 379

RESULT 11  
 US-09-077-991-2  
 Sequence 2, Application US/09077991  
 Patent No. 6207375  
 GENERAL INFORMATION:  
 APPLICANT: Sudramaniam, M.  
 APPLICANT: Spelsberg, T.C.  
 APPLICANT: Roche, P.C.  
 TITLE OF INVENTION: TGF-Beta Inducible early factor-1  
 FILE REFERENCE: (TIEF-1) and a method to detect breast cancer  
 CURRENT APPLICATION NUMBER: US/09/077,991  
 EARLIER APPLICATION NUMBER: PCT/US96/19555  
 EARLIER FILING DATE: 1996-12-11  
 EARLIER APPLICATION NUMBER: US 08/570,227  
 NUMBER OF SEQ ID NOS: 13  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 2  
 LENGTH: 480  
 TYPE: PRT  
 ORGANISM: Homo sapiens  
 US-09-077-991-2

Alignment Scores:  
 Pred. No.: 1,49  
 Score: 78.00  
 Percent Similarity: 33.57%  
 Best Local Similarity: 23.21%  
 Query Match: 5.24%  
 DB: 4  
 Gaps: 14

US-09-763-902B-10 (1-830) x US-09-077-991-2 (1-480)

```

OY 11 CATTACGCGGCTCTCTTCAACCTTACTCC-----TCAGCCCGCTGGTCCCAAG 64
DB 118 Hispelyserleuserprhralalysprohlsilaalalapro--Pheylsglu 136
OY 65 GCAGAGAAATACCGCTGGAGCCCTCTTCATGAAACGACATGTCATGTGGAT 124
DB 137 Glugluysserprovalserlarpolylsleuprolysalaglnlathrservalile 156
OY 125 TTCAGACCTATGTGATCAGCGCTGCAGA-----GCTGCTAGAGAG 166
DB 157 ArgHsthrAlaAspAlaglnleucysasnHsglnthrcysPrometlyslalaaser 176
OY 167 TTTCGCAATGTCTACTACACCATGATAGCGGCGGCTTG-----CTGTCC 217
DB 177 IleleuasnTyrglnAsnAsnSerPheargTargTgthHlsleuAsnValgluAlaala 196
OY 218 CGCGTGTACGGGACAGCCACCTGCTGTGATGCAATGCT-----GTCGCTAGAGAG 166
DB 197 ArglyAsnIleProcyAlaAlaValSerProAsnArgserlyscysgluArgasnThr 216
OY 263 -----GTTTCAGACAGAAATCCTTGAGTGAAGTTTGGTAATGTGCTCCAGC 313
DB 217 ValAlaAspValaAspGlnlyslaserAlaAlaLeuTyrgAspPheSerValaProserSer 236
OY 263 -----GTTTCAGACAGAAATCCTTGAGTGAAGTTTGGTAATGTGCTCCAGC 313
DB 217 ValAlaAspValaAspGlnlyslaserAlaAlaLeuTyrgAspPheSerValProserSer 236

```

OY 314 GAGTTCGAATACAGCGGTAGACAGCCAGCTGTTCATGATGAAGACCAACAGCCAG 373  
 DB 237 Gln-----ThValIleCysArgSerGlnProAlaProValSerProGlnIn 252  
 OY 374 ACCAGGCTCTGTT-----GTC 391  
 DB 253 LysSerValIleValSerProProAlaValSerAlaGlyGlyValProProMetProVal 272  
 OY 392 ATGTGTGATCAGTGAAGTGTGAGGAGCAACACAGGAGCTTACACAGACTTCATC 451  
 DB 273 IleCysGlnMetValProLeuProAlaAsn-AsnProValValThrThrValValProSe 292  
 OY 452 CTGACCCGCCAG-----464  
 DB 292 rThProProSerGlnProProAlaValCysProProValValPheMetGlyThGlnVa 312  
 OY 465 -----CCTCACCCAGCAGACAGCTGTGAGATCGCA 496  
 DB 312 lProLysGlyAlaValMetPheValValProGlnProValValGlnSerSerLysProPr 332  
 OY 497 AGTCACTGCT-----TCCGCTTCAGAGCTGGCCAGCTAGTGGGGGTGGCAG 544  
 DB 332 oValValSerProAsnGlyThrArgLeuSerProIleAlaProAlaProGly-----349  
 OY 545 AGGTCTCTTCTTATTCAGCCCTAGCTCTAGAGAAATGCMAACTCGACTC-----599  
 DB 350 -----PheSerProSerAlaAla--LysValThProGlnIleAspSe 363  
 OY 600 -TCAGAGATGTAGAGAACAGTTCATTTCTGTTGT-----TGGGAGACACT 647  
 DB 363 rSerArgIle-----ArgSerHisIleCysSerHisProGlyCysGlyLysThr 379

RESULT 12  
 US-08-290-301-82  
 Sequence 82, Application US/08290301  
 Patent No. 5792921  
 GENERAL INFORMATION:  
 APPLICANT: Lonsdesborough, John  
 APPLICANT: Tunnela, Outi  
 APPLICANT: Palva, Tupio  
 APPLICANT: Holmstrom, Kjell-Ove  
 APPLICANT: Mandel, Abul  
 TITLE OF INVENTION: Increasing the trehalose content  
 TITLE OF INVENTION: of organisms by transforming them with combinations of  
 TITLE OF INVENTION: the structural genes for trehalose synthase.  
 NUMBER OF SEQUENCES: 85  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Alko Ltd.  
 STREET: PO Box 350  
 CITY: Helsinki  
 STATE: Finland  
 COUNTRY: Finland  
 ZIP: SF-00101  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb  
 COMPUTER: IBM PC/XT/AT  
 OPERATING SYSTEM: PC-DOS  
 SOFTWARE: WP5.1 file exported as DOS text file  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/290, 301  
 FILING DATE: 15 August 1994  
 CLASSIFICATION: 424  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: FI 943133  
 FILING DATE: 29 June 1994  
 APPLICATION NUMBER: PCT/FI93/00049  
 FILING DATE: 15 February 1993  
 APPLICATION NUMBER: 07/841,997  
 FILING DATE: 28 February 1992  
 APPLICATION NUMBER: 07/836,021  
 FILING DATE: 14 February 1992  
 ATTORNEY/AGENT INFORMATION:

NAME: Kubovcik, Ronald J.  
 NAME: Lydon, James C.  
 REGISTRATION NUMBER: 25,401  
 REGISTRATION NUMBER: 30,082  
 REFERENCE/DOCKET NUMBER: LAIIN-001  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 467-6300  
 TELEFAX: (202) 466-2006  
 INFORMATION FOR SEQ ID NO: 82:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1098 amino acids  
 TYPE: Amino acid  
 TOPOLOGY: Linear  
 MOLECULE TYPE: Polypeptide  
 HYPOTHETICAL: Yes  
 US-08-290-301-82

Alignment Scores:  
 Pred. No.: 2,3 Length: 1098  
 Score: 78.00 Matches: 27  
 Percent Similarity: 47.37% Conservative: 9  
 Best Local Similarity: 35.53% Mismatches: 28  
 Query Match: 5.24% Indels: 12  
 Gaps: 3

US-09-763-902b-10 (1-830) x US-08-290-301-82 (1-1098)  
 OY 381 TCCTTTTTCATCCTGTGATCAGTGTGAGGGAACAAACAGGAGCTTCAACC 440  
 DB 186 SerLeuLeuAsnHisThrSerGlnThrSerLeuGlyGlyProAsnHisIleValThr 205  
 OY 441 AGAATTCATCCTCGAGCCGCCAGCTTCACCCAGCAACAGTGTGAAGATCGCAAGTC 500  
 DB 206 ProLysSerArgAlaGlyAsnArgPro-----ThrSerAlaAla 218  
 OY 501 ACTGCTTCGCTTCAGAGACT-----GGCCAGCTAGTGGGGGTGGCAGAGTCTCTT 554  
 DB 219 ThrSerLeuValAsnArgThrLysGlnGlySerAlaSerSerLysSerGly-----236  
 OY 555 GCTTCATTCAGCCCTAGCTCTGTAGAGAAATGCMAACTCGACTCTCA 602  
 DB 237 ---SerSerAlaProProSerIleLysArgIleThrProHisLeuThr 251

RESULT 13  
 US-09-013-598-82  
 Sequence 82, Application US/09013598  
 Patent No. 6323001  
 GENERAL INFORMATION:  
 APPLICANT: Lonsdesborough, John  
 APPLICANT: Tunnela, Outi  
 APPLICANT: Palva, Tupio  
 APPLICANT: Holmstrom, Kjell-Ove  
 APPLICANT: Mandel, Abul  
 TITLE OF INVENTION: Increasing the trehalose content  
 TITLE OF INVENTION: of organisms by transforming them with combinations of  
 TITLE OF INVENTION: the structural genes for trehalose synthase.  
 NUMBER OF SEQUENCES: 85  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Alko Ltd.  
 STREET: PO Box 350  
 CITY: Helsinki  
 STATE: Finland  
 COUNTRY: Finland  
 ZIP: SF-00101  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Diskette, 3.5 inch, 720 Kb  
 COMPUTER: IBM PC/XT/AT  
 OPERATING SYSTEM: PC-DOS  
 SOFTWARE: WP5.1 file exported as DOS text file  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/09/013, 598  
 FILING DATE:

CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/290,301  
FILING DATE:  
APPLICATION NUMBER: PCT/E193/00049  
FILING DATE: 15 February 1993  
APPLICATION NUMBER: 07/841,997  
FILING DATE: 28 February 1992  
APPLICATION NUMBER: 07/836,021  
FILING DATE: 14 February 1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Kubovcik, Ronald J.  
REGISTRATION NUMBER: 25,401  
REFERENCE/DOCKET NUMBER: 30,082  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 467-6300  
TELEFAX: (202) 466-2006  
INFORMATION FOR SEQ ID NO: 82:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1098 amino acids  
TYPE: Amino acid  
TOPOLOGY: Linear  
MOLECULE TYPE: Polypeptide  
HYPOHETICAL: yes  
US-09-013-598-82

Alignment Scores:  
Pred. No.: 2,3  
Score: 78.00  
Percent Similarity: 47.37%  
Best Local Similarity: 35.53%  
Query Match: 5.24%  
DB: 4  
Length: 1098  
Matches: 27  
Conservative: 9  
Mismatch: 28  
Indels: 12  
Gaps: 3

US-09-763-902b-10 (1-830) x US-09-013-598-82 (1-1098)

QY 381 TCGTGTCTGTCATGTCGTGATGATGAGTTCAGAGGAGCAACAGCGACTTCACAC 440  
DB 186 SerLeuLeuAsnAsnThrSerLeuThrSerLeuGluGlyProAsnAsnHisLeuAlaThr 205  
QY 441 AGAATTCATCTGAGCCGCCAGGCCCTCAGCCAGACACAGTGTGAAGATGCGAAGT 500  
DB 206 ProLysSerArgAlaGlyAsnArgPro-----ThSerAlaAla 218  
QY 501 ACTGCTTCGCTCCAGACT-----GGCCAGCTAGTGGGGGTGGCAGAGTCTCTT 554  
DB 219 ThrSerLeuValAsnArgThrLysGlnGlySerAlaSerGlySerGly----- 236  
QY 555 GCTTCATTCAGCCCTGAGCTGTGAGAGAAATGCAAACTGCACTCTCA 602  
DB 237 ---SerSerAlaProProSerLeuLysArgGlyLeuThrProHisLeuThr 251

RESULT 14  
US-08-216-260-4  
Sequence 4, Application US/08216260  
Patent No. 5837261  
GENERAL INFORMATION:  
APPLICANT: Ingllis, Stephen C.  
APPLICANT: Bousnell, Michael E. G.  
APPLICANT: Minson, Anthony C.  
TITLE OF INVENTION: VIRAL VACCINES  
NUMBER OF SEQUENCES: 12  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Flehr, Hohbach, Test, Albritton & Herbert  
ADDRESSEE: Attn: Walter H. Dreyer  
STREET: 4 Embarcadero Center, Suite 3400  
CITY: San Francisco  
STATE: California  
COUNTRY: United States  
ZIP: 94111-4187  
COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/216,260  
FILING DATE: 21-MAR-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9305710.7  
FILING DATE: 19-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: GB 9324964.7  
FILING DATE: 06-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/030,073  
FILING DATE: 19-MAR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/168,643  
FILING DATE: 16-DEC-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreyer, Walter H.  
REGISTRATION NUMBER: 24,190  
REFERENCE/DOCKET NUMBER: A-59886/WHD  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (415) 781-1989  
TELEFAX: (415) 398-3249  
TELEX: 910 277299  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 838 amino acids  
TYPE: amino acid  
STRANDEDNESS: both  
TOPOLOGY: linear  
US-08-216-260-4

Alignment Scores:  
Pred. No.: 2,64  
Score: 77.00  
Percent Similarity: 35.45%  
Best Local Similarity: 25.45%  
Query Match: 5.17%  
DB: 2  
Length: 838  
Matches: 56  
Conservative: 22  
Mismatch: 66  
Indels: 76  
Gaps: 13

US-09-763-902b-10 (1-830) x US-08-216-260-4 (1-838)

QY 51 CTTGTTCCCAAGCAGAGAAATACCTGTCGAGCCCTTCATAGACAGAGA 110  
DB 53 ProAsnThrProAspProGlnLysProProArgGlyPheLeuAlaPro----- 68  
QY 111 TGGCATCTGTGATTTCAAGACCTATGTGATCAGGCCCTGAGAGCTGAGAGTTG 170  
DB 69 -----ProAspGlnLeu-----AsnLeu 74  
QY 171 TCAATGCTACTACCA-----CCATGATTAACGGCGCTTGTCTCCGCTGT 224  
DB 75 ThrThrAlaSerLeuProLeuLeuArgTrpGluGluArgPheCysPheValLeuVal 94  
QY 225 ACATGGCAGACCCAGCCCTGCTGCAATGCAATGCTGTTTCAGCAAGAAAT----- 278  
DB 95 ThrThrAlaGluPheProArgAspProGlyGlnLeuLeuTrpIleProLysThrTyLeu 114  
QY 278 ----- 278  
DB 115 LeuGlyArgProProAsnAlaSerLeuProAlaProThrThrValGluProThrAlaGln 134  
QY 279 -----CCTTGAAGATTTTGAATGTCCTT-----CCAGGAATTC----- 320  
DB 135 ProProProSerValAlaProLeuLysGlyLeuLeuHisAsnProAlaAlaSerValLeu 154  
QY 321 -----AAATCAGCGTGGAGACTGCCAGCCCTGTCATGATGAAGCCACACAGCAGCA 374  
DB 155 LeuArgSerArgAlaTrpValThr-----PheSerAlaValProAsp----- 168

Search completed: June 24, 2003, 19:40:14  
Job time: 23.711 secs

```
QY 375 CCACGGTCCTTTGTCATCTGTGATCAGTGAAGTTGAGGGGAAACAACGGGACT 434
    ||| :|||
Db 169 ProGluAlaLeuThPhePro-----ArgGlyAspAsnValAlaThr 182
QY 435 TCACACGACAACTTCA-----TCCTGACCGCCCGAGCCCT----- 467
    :||| :|||
Db 183 AlaSerHisProSerGlyProArgAspThrProProArgProProValGlyAlaArg 202
QY 468 ---CACCCAGACACACAGTGTGAGAGATGCAAGTACTGCTCCGCTTCACAGACTGGG 524
    ||||| :|||
Db 203 ArgHisProThrThGluLeuAspIleThrHisLeu---HisAsnAlaSerThrThrTrp 221
QY 525 CCAGCTAGTGGGGGT-----GGCAGAGGCTCTTCTTCATTCACGCCCTACTGCT 575
    ||||| :|||
Db 222 LeuAlaThrArgGlyLeuLeuArgSerProGlyArgGlyValIlePheSerProSerAla 241
```

## RESULT 15

```
US-08-846-762-5
: Sequence 5, Application US/08846762A
: Patent No. 5994072
:
: GENERAL INFORMATION:
: APPLICANT: Lam, Joseph S.
: APPLICANT: Burrows, Lori
: APPLICANT: Charter, Deborah
: APPLICANT: de Kievit, Teresa
:
: TITLE OF INVENTION: No. 5994072e1 Proteins Involved in the Synthesis and Assembly
: FILE REFERENCE: 6580-089
: CURRENT APPLICATION NUMBER: US/08/846,762A
: CURRENT FILING DATE: 1997-04-30
: NUMBER OF SEQ ID NOS: 100
: SOFTWARE: Patentln Ver. 2.0
:
: SEQ ID NO 5
: LENGTH: 766
: TYPE: PRT
: ORGANISM: Pseudomonas aeruginosa
US-08-846-762-5
```

## Alignment Scores:

```
Pred. No.: 2.89 Length: 766
Score: 76.50 Matches: 28
Percent Similarity: 39.66 Conservative: 18
Best Local Similarity: 24.14 Mismatches: 55
Query Match: 5.14 Indels: 15
DB: 2 Gaps: 3
```

US-09-763-902B-10 (1-830) x US-08-846-762-5 (1-766)

```
QY 59 CCCAAGGACAGAAATACCTGTGAGAGCCCTCCCTTCCATAGACCAAGATGCATCT 118
    |||:|||||
Db 277 ProArgAlaGluLeuLeuProLeuLeuHisThrCysSerIleAlaAsnGluMet----- 294
QY 119 GTGGATTTCAGACACTATGTGATCAGGCTCGAGAGCTGCTGAGAGTTTGTCAATGTC 178
    :||| :|||
Db 295 -----GlnPheThrLeu-----PheTyrProVal 302
QY 179 TACTACACCAACATGATAGCGCGCGCTTGTGCTGTCCGCTGTACATGAGGCACAGCC 238
    :||| :|||
Db 303 LeuPheMetCysLeuProCysArgTrpArgLeuProValPheIleLeuAlaIleLeu 322
QY 239 ACCCTGTCTGTGAATGCAATGCTGTT---TCAGACACAGAAATCCTTGAGTGAATTGTTT 295
    :||| :|||
Db 323 LeuPheIleTrpSerIleTyrCysValPheSerGlySerGlnAspAlaGlnTyrPheAla 342
QY 296 GAAATGTTGCTTCAGCGAGTTCACAAATCAGCGTGTAGACGCCAGCCTGTTCATGAT 355
    :||| :|||
Db 343 LeuLeuAlaArgValProGluPheMetSerGlyAlaValAlaLeuSerLeuArgAsp 362
QY 356 GAAGCCACACACAGCAGCAGCTCTGTTGTGATCTGTGATCA 403
    |||:|||||
Db 363 ArgGluLeuProAlaArgLeuAlaIleLeuAlaGlyLeuLeuGlyAla 378
```

GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 24, 2003, 19:31:14 ; Search time 31.503 Seconds

(Without alignments)  
5701.778 Million cell updates/sec

Title: US-09-763-902B-10

Perfect score: 1489  
Sequence: 1 tgaacttgcatcactgctg.....ataattaataacacatg 830

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 417779 segs, 108206813 residues

Total number of hits satisfying chosen parameters: 835558

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame\_plus\_n2p.model -DEV=xlp  
-O=/cgn2.1/USPTO/US09/63902/runat\_24062003\_130343\_12522/app\_query.fasta.1.2254  
-DB=PublishedApplications\_AA -QFMT=FASTA -SUFFIX=n2p.rapb -MINMATCH=0.1  
-LOOCL=0 -LOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsu62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MUTLEN=0  
-MAXLEN=200000000 -USER=US09763902.ecgn\_1\_1\_24@runat\_24062003\_130343\_12522  
-NCPU=6 -ICPU=3 -NO\_MAP -LARGEQUERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100  
-LONLOG -DEV=TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : PublishedApplications\_AA.\*

1: /cgn2.6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
2: /cgn2.6/ptodata/2/pubpaa/FCI\_NEW\_PUB.pep.\*  
3: /cgn2.6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2.6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2.6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2.6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
7: /cgn2.6/ptodata/2/pubpaa/FCIS\_PUBCOMB.pep.\*  
8: /cgn2.6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2.6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep.\*  
10: /cgn2.6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2.6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
12: /cgn2.6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
13: /cgn2.6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
14: /cgn2.6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	813	54.6	164	US-09-764-864-1181
2	579	38.9	10	Sequence 1181, Ap
3	115.5	7.8	449	US-09-764-864-1180
4	105.5	7.1	9	Sequence 272, App
				Sequence 318, App

	5	105.5	7.1	477	10	US-09-925-301-977	Sequence 977, App
	6	89	6.0	326	10 <td>US-09-764-864-1044</td> <td>Sequence 1044, Ap</td>	US-09-764-864-1044	Sequence 1044, Ap
	7	86.5	5.8	462	10 <td>US-09-814-777A-99</td> <td>Sequence 99, Appl</td>	US-09-814-777A-99	Sequence 99, Appl
	8	85.5	5.7	802	10 <td>US-09-907-495-2</td> <td>Sequence 2, Appl1</td>	US-09-907-495-2	Sequence 2, Appl1
	9	85.5	5.7	1604	10 <td>US-09-746-491-8</td> <td>Sequence 8, Appl1</td>	US-09-746-491-8	Sequence 8, Appl1
	10	83.5	5.6	197	9 <td>US-10-149-328-8</td> <td>Sequence 8, Appl1</td>	US-10-149-328-8	Sequence 8, Appl1
	11	83.5	5.6	475	10 <td>US-09-741-669-458</td> <td>Sequence 458, App</td>	US-09-741-669-458	Sequence 458, App
	12	80.5	5.4	831	9 <td>US-10-273-680-2</td> <td>Sequence 2, Appl1</td>	US-10-273-680-2	Sequence 2, Appl1
	13	80	5.4	1276	10 <td>US-09-982-610-24</td> <td>Sequence 24, Appl1</td>	US-09-982-610-24	Sequence 24, Appl1
	14	78.5	5.3	725	9 <td>US-10-160-501-17</td> <td>Sequence 17, Appl1</td>	US-10-160-501-17	Sequence 17, Appl1
	15	78.5	5.3	2971	9 <td>US-10-146-472-50</td> <td>Sequence 50, Appl1</td>	US-10-146-472-50	Sequence 50, Appl1
	16	77.5	5.2	221	9 <td>US-10-102-806-603</td> <td>Sequence 603, Appl1</td>	US-10-102-806-603	Sequence 603, Appl1
	17	77.5	5.2	1057	9 <td>US-10-189-971-6</td> <td>Sequence 6, Appl1</td>	US-10-189-971-6	Sequence 6, Appl1
	18	77.5	5.2	1251	9 <td>US-10-189-971-16</td> <td>Sequence 16, Appl1</td>	US-10-189-971-16	Sequence 16, Appl1
	19	77.5	5.2	1342	9 <td>US-10-189-971-24</td> <td>Sequence 24, Appl1</td>	US-10-189-971-24	Sequence 24, Appl1
	20	77.5	5.2	1512	9 <td>US-10-189-971-10</td> <td>Sequence 10, Appl1</td>	US-10-189-971-10	Sequence 10, Appl1
	21	77.5	5.2	1570	9 <td>US-10-189-971-12</td> <td>Sequence 12, Appl1</td>	US-10-189-971-12	Sequence 12, Appl1
	22	77.5	5.2	1628	9 <td>US-10-189-971-2</td> <td>Sequence 2, Appl1</td>	US-10-189-971-2	Sequence 2, Appl1
	23	76	5.1	190	9 <td>US-09-738-626-6057</td> <td>Sequence 6057, Ap</td>	US-09-738-626-6057	Sequence 6057, Ap
	24	76	5.1	398	9 <td>US-10-097-340-248</td> <td>Sequence 248, App</td>	US-10-097-340-248	Sequence 248, App
	25	76	5.1	1106	9 <td>US-10-157-031-30</td> <td>Sequence 30, Appl1</td>	US-10-157-031-30	Sequence 30, Appl1
	26	76	5.1	2330	9 <td>US-10-184-644-155</td> <td>Sequence 155, App</td>	US-10-184-644-155	Sequence 155, App
	27	76	5.1	2530	9 <td>US-10-184-634-155</td> <td>Sequence 155, App</td>	US-10-184-634-155	Sequence 155, App
	28	75.5	5.1	242	10 <td>US-09-764-853-609</td> <td>Sequence 609, App</td>	US-09-764-853-609	Sequence 609, App
	29	75.5	5.1	242	10 <td>US-09-764-898-199</td> <td>Sequence 199, App</td>	US-09-764-898-199	Sequence 199, App
	30	75.5	5.1	619	9 <td>US-09-976-059-30</td> <td>Sequence 30, Appl1</td>	US-09-976-059-30	Sequence 30, Appl1
	31	75.5	5.1	785	10 <td>US-09-801-368-348</td> <td>Sequence 348, App</td>	US-09-801-368-348	Sequence 348, App
	32	75.5	5.1	963	9 <td>US-10-140-164-2</td> <td>Sequence 2, Appl1</td>	US-10-140-164-2	Sequence 2, Appl1
	33	75.5	5.1	1027	9 <td>US-10-140-164-4</td> <td>Sequence 4, Appl1</td>	US-10-140-164-4	Sequence 4, Appl1
	34	75.5	5.1	2871	9 <td>US-10-146-473-41</td> <td>Sequence 41, Appl1</td>	US-10-146-473-41	Sequence 41, Appl1
	35	75	5.0	316	9 <td>US-10-153-668-134</td> <td>Sequence 134, App</td>	US-10-153-668-134	Sequence 134, App
	36	75	5.0	323	9 <td>US-10-153-668-132</td> <td>Sequence 132, App</td>	US-10-153-668-132	Sequence 132, App
	37	75	5.0	370	9 <td>US-10-153-668-130</td> <td>Sequence 130, App</td>	US-10-153-668-130	Sequence 130, App
	38	75	5.0	410	9 <td>US-09-870-753-48</td> <td>Sequence 48, Appl1</td>	US-09-870-753-48	Sequence 48, Appl1
	39	75	5.0	473	9 <td>US-09-978-295A-400</td> <td>Sequence 400, App</td>	US-09-978-295A-400	Sequence 400, App
	40	75	5.0	473	9 <td>US-09-978-697-400</td> <td>Sequence 400, App</td>	US-09-978-697-400	Sequence 400, App
	41	75	5.0	473	9 <td>US-09-978-192A-400</td> <td>Sequence 400, App</td>	US-09-978-192A-400	Sequence 400, App
	42	75	5.0	473	9 <td>US-09-999-832A-400</td> <td>Sequence 400, App</td>	US-09-999-832A-400	Sequence 400, App
	43	75	5.0	473	9 <td>US-09-978-189-400</td> <td>Sequence 400, App</td>	US-09-978-189-400	Sequence 400, App
	44	75	5.0	473	9 <td>US-10-028-072-382</td> <td>Sequence 382, App</td>	US-10-028-072-382	Sequence 382, App
	45	75	5.0	473	9 <td>US-10-121-049-382</td> <td>Sequence 382, App</td>	US-10-121-049-382	Sequence 382, App

#### ALIGNMENTS

RESULT 1  
US-09-764-864-1181  
Sequence 1181, Application US/09764864  
Patent No. US20020132753A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
FILE REFERENCE: PT223  
CURRENT APPLICATION NUMBER: US/09/764,864  
CURRENT FILING DATE: 2001-01-17  
Prior application data removed - consult PALM or file wrapper  
NUMBER OF SEQ ID NOS: 1792  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 1181  
LENGTH: 164  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (2)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (3)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (5)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE

LOCATION: (10)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-764-864-1181

**Alignment Scores:**

Pred. No.:	2.81e-77	Length:	164
Score:	813.00	Matches:	153
Percent Similarity:	100.00%	Conservative:	0
Best Local Similarity:	100.00%	Mismatches:	0
Query Match:	54.60%	Indels:	0
DB:	10	Gaps:	0

US-09-763-902B-10 (1-830) x US-09-764-864-1181 (1-164)

OY	71	GAATATACCTGGGAGAGCCCTCTTCATATAGAACAGAGATGGCATGTGGATTTCAG	130
Dd	12	GIUIIeProTIPtPseRProPseRIIeGIuPogIuMeAlaSerValAspPheLys	31
OY	131	ACCTATGTGAATCAGAGCCCTGCAGAGACCTGCTGAGAGATTTCATATGCTACTACACACC	190
Dd	32	ThItyValAspInAlaIeLacysArgAlaIaGIuInIuPheValAsnValTytTytThr	51
OY	191	ATGGATTAACGGGGGGGTTGCTGCTGTCGCCCTGTATCATATGGGACACAGCCACCTGTGTGG	250
Dd	52	MeAspLysArgTArgTArgIeuLeuSerArgIeuTyrMetGIyThrAlaThrIeuValTyr	71
OY	251	AATGGCAATGCTGATTTCAGAGACAAGAATCCTTGATGATGATTTTGAATTTGCTTCC	310
Dd	72	AsnGIYAsnAlaValSerGIYngIuInSerIeuSerGIuPheNegIuMetIeuProSer	91
OY	311	AGGAGATTCCAAATCAAGCCTGGTATACAGCCAGCCTGTTCATGATGAAGCACACCAAGC	370
Dd	92	SerGIuPheInIleSerValValAspGlyInPheValHisAspGIuAlaThrProSer	111
OY	371	CAGACCAAGCTCCTTTGTCATCTGTGATCATGAATTCAGAGGGGACAAACAACAGG	430
Dd	112	GIuThrThrValIeuValIValIleCysGIySerValLysPheGIuGIYAsnLysGIuArg	131
OY	431	GACTTCAACACAGACTTCATCTCTGACCGCCACAGGCTCAACCAAGACACAGTGTGAG	490
Dd	132	AspPheAsnGIuInAsnPheIleIeuThrAlaGIuAlaIaSerProSerAsnThrValTyrLys	151
OY	491	ATGCAAGTACATGCTTCGGCTTCAAGACATGGGGCAGC	529
Dd	152	IleIaSerAspCysPheArgPheGIuAspTrrPalaser	164

## RESULT 2

```

US-09-764-864-1180
: Sequence 1180, Application US/09764864
: Patent No. US2002012753A1
: GENERAL INFORMATION:
: APPLICANT: Rosen et al.
: TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
: FILE REFERENCE: PT623
: CURRENT APPLICATION NUMBER: US/09/764, 864
: CURRENT FILING DATE: 2001-01-17
: Prior application data removed - consult PALM or file wrapper
: NUMBER OF SEQ ID NOS: 1792
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 1180
: LENGTH: 162
: TYPE: PRT
: ORGANISM: Homo sapiens
US-09-764-864-1180

```

**Alignment Scores:**

Pred. No.:	1.3e-52	Length:	165
Score:	579.00	Matches:	10
Percent Similarity:	91.18%	Conservative:	21
Best Local Similarity:	75.74%	Mismatches:	12
Query Match:	38.89%	Indels:	0
DB:	10	Gaps:	0

US-09-763-902B-10 (1-830) X US-09-764-864-1180 (1-162)

[illegible]

### RESULT 3

```

US-10-074-475-272
: Sequence 272, Application US/10074475
: Publication No. US20030092898A1
:
: GENERAL INFORMATION:
: APPLICANT: Salceda, Susana
: APPLICANT: Macina, Roberto
: APPLICANT: Hu, Ping
: APPLICANT: Recipon, Hervé
: APPLICANT: Katta, Kalpana
: APPLICANT: Caffeekey, Robert
: APPLICANT: Sun, Yongming
: APPLICANT: Liu, Chenghua
:
: TITLE OF INVENTION: Compositions and Methods Relating to Breast Specific
: FILE OF INVENTION: Genes and Proteins
: FILE REFERENCE: DEX-0313
:
: CURRENT APPLICATION NUMBER: US/10/074,475
:
: CURRENT FILING DATE: 2002-02-13
:
: PRIOR APPLICATION NUMBER: 60/268,292
:
: PRIOR FILING DATE: 2001-02-13
:
: NUMBER OF SEQ ID NOS: 295
:
: SOFTWARE: PatentIn version 3.1
:
: SEQ ID NO 272
:
: LENGTH: 449
:
: TYPE: prt
:
: ORGANISM: Homo sapien
:
US-10-074-475-272

```

**Alignment Scores:**

Pred. No.:	0.00273	Length:	44
Score:	115.50	Matches:	38
Percent Similarity:	44.09%	Conservative:	18
Best Local Similarity:	29.92%	Mismatches:	58
Query Match:	7.76%	Indels:	13
Ds:	9	Gaps:	4

US-09-763-902B-10 (1-830) x US-10-074-475-272 (1-449)

QY 164 GAGTTTGTCAAAGTCTACTACACCAACGATAGAGGGGGGCTTGGCTGTCCGGCTG 22

||||||| ||||| ||| |||

Db 14 GuphevalaArgSlnTyrtYrThrIleuAspLysAlaIaProGlnTyrtLeuHISArgPhe 33

||||||| ||||| ||||| |||||

QY 224 TACATGGGACGACGCCACCTCGTGTCTGGAATGGC-----AAT 255

DB 34 TyrglyArgasnSerSerTyValHisglYlyValAspAlaSerGlyLysProGlnGlu 53  
QY 260 GCTGTTTACGACAGAAATCCCTGAGTGTGTTTGAATGTCCTCCAGCGACTC 319  
DB 54 AlaValTyrglyGlnAsnAspIleHisLysValLeuSerLeuAsnPheserGlyLys 73  
QY 320 CAATACAGCGGTGAGTGGCCGCTGTCATGATGAAGCCACCAAGCCAGACGACG 379  
DB 74 HisThrLysIle-----ArgHisValAspAlaHisAlaThrLeuSerAsp---Gly 89  
QY 380 GTCCCTTGTTCATCTGTGATCATGATCAAGTTTGAGGGAACAACAGCGACTTCAC 439  
DB 90 ValValValGlnValMetGlyLeuLeuSerAsnSerGlyGlnProGlnArgLysPheMet 109  
QY 440 CAGAACTTCATCCGTGACCGCCCGCCCTCCACCAACACAGTGTGAAGTCCGACT 499  
DB 110 GlnThrPheValLeuAlaProGlnLysSerValProAsnLysPheTy---ValHisAsn 128  
QY 500 GACTGCTTCGCTTCAGGAC 520  
DB 129 AspMetPheArgTyGlnAsp 135

RESULT 4  
US-09-919-039-318  
Sequence 318, Application US/09919039  
Patent No. US20030108871A1  
GENERAL INFORMATION:  
APPLICANT: Kaser, Matthew R.  
TITLE OF INVENTION: GENES EXPRESSED IN TREATED HUMAN C3A LIVER CELL CULTURES  
FILE REFERENCE: PA-0035 US  
CURRENT APPLICATION NUMBER: US/09/919, 039  
CURRENT FILING DATE: 2002-09-09  
PRIOR APPLICATION NUMBER: 60/222,113  
PRIOR FILING DATE: 2000-07-28  
NUMBER OF SEQ ID NOS: 401  
SOFTWARE: PERL Program  
SEQ ID NO 318  
LENGTH: 466  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc\_feature  
OTHER INFORMATION: Incyte ID No. US20030108871A1 2477616CD1  
US-09-919-039-318

Alignment Scores:  
Pred. No.: 0.0317 Length: 466  
Score: 105.50 Matches: 36  
Percent Similarity: 44.88% Conservative: 21  
Best Local Similarity: 28.35% Mismatches: 57  
Query Match: 7.09% Indels: 13  
DB: 9 Gaps: 4

US-09-763-902b-10 (1-830) x US-09-919-039-318 (1-466)  
QY 164 GAGTTTGTCAATGTCTACTACACCAACGATGAAGCGGGCGTTGTGCTCCGCTG 223  
DB 14 GluPheValArgLysIleThrLeuLeuAsnGlnAlaProAspMetLeuHisArgPhe 33  
QY 224 TACATGGGACAGCCACCCCTGCTGTGATGGC-----AAT 259  
DB 34 TyrglyLysAsnSerSerTyValHisglYlyLeuAspSerAsnGlyLysProAlaAsp 53  
QY 260 GCGTTTCAAGCAAGAAATCCCTGAGTGTGTTTGAATGTCCTCCAGCGACTTC 319  
DB 54 AlaValTyrglyGlnLys-----GlnIleHisArgLysValMetSerGlnAsnPh 70  
QY 320 CAATACAGCGGTGAGTGGCCGCTGTCATGATGAAGCCACCAAGCCAGACGACG 379  
DB 71 ThrAsnGlyHisThrLysIleArgHisValAspAlaHisAlaThrLeuAsnAsp---Gly 89  
QY 380 GTCCCTTGTTCATCTGTGATCATGATCAAGTTTGAGGGAACAACAGCGACTTCAC 439

DB 90 ValValValGlnValMetGlyLeuLeuSerAsnAsnGlnAlaLeuArgArgPheMet 109  
QY 440 CAGAACTTCATCCGTGACCGCCCGCCCTCCACCAACACAGTGTGAAGTCCGACT 499  
DB 110 GlnThrPheValLeuAlaProGlnLysSerValAlaHisLysPheTy---ValHisAsn 128  
QY 500 GACTGCTTCGCTTCAGGAC 520  
DB 129 AspIlePheArgTyGlnAsp 135

RESULT 5  
US-09-925-301-977  
Sequence 977, Application US/09925301  
Patent No. US20020052308A1  
GENERAL INFORMATION:  
APPLICANT: Rosen et al.  
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies  
FILE REFERENCE: PA106  
CURRENT APPLICATION NUMBER: US/09/925, 301  
CURRENT FILING DATE: 2001-08-10  
PRIOR APPLICATION NUMBER: PCT/US00/05882  
PRIOR FILING DATE: 2000-03-08  
PRIOR APPLICATION NUMBER: 60/124, 270  
PRIOR FILING DATE: 1999-03-12  
NUMBER OF SEQ ID NOS: 1694  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 977  
LENGTH: 477  
TYPE: PRT  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: SITE  
LOCATION: (471)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
NAME/KEY: SITE  
LOCATION: (473)  
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids  
US-09-925-301-977

Alignment Scores:  
Pred. No.: 0.032 Length: 477  
Score: 105.50 Matches: 36  
Percent Similarity: 44.88% Conservative: 21  
Best Local Similarity: 28.35% Mismatches: 57  
Query Match: 7.09% Indels: 13  
DB: 10 Gaps: 4

US-09-763-902b-10 (1-830) x US-09-925-301-977 (1-477)  
QY 164 GAGTTTGTCAATGTCTACTACACCAACGATGAAGCGGGCGTTGTGCTCCGCTG 223  
DB 25 GluPheValArgLysIleThrLeuLeuAsnGlnAlaProAspMetLeuHisArgPhe 44  
QY 224 TACATGGGACAGCCACCCCTGCTGTGATGGC-----AAT 259  
DB 45 TyrglyLysAsnSerSerTyValHisglYlyLeuAspSerAsnGlyLysProAlaAsp 64  
QY 260 GCTGTTTCAAGCAAGAAATCCCTGAGTGTGTTTGAATGTCCTCCAGCGACTTC 319  
DB 65 AlaValTyrglyGlnLys-----GlnIleHisArgLysValMetSerGlnAsnPh 81  
QY 320 CAATACAGCGGTGAGTGGCCGCTGTCATGATGAAGCCACCAAGCCAGACGACG 379  
DB 82 ThrAsnGlyHisThrLysIleArgHisValAspAlaHisAlaThrLeuAsnAsp---Gly 100  
QY 380 GTCCCTTGTTCATCTGTGATCATGATCAAGTTTGAGGGAACAACAGCGACTTCAC 439  
DB 101 ValValValGlnValMetGlyLeuLeuSerAsnAsnGlnAlaLeuArgArgPheMet 120  
QY 440 CAGAACTTCATCCGTGACCGCCCGCCCTCCACCAACACAGTGTGAAGTCCGACT 499  
DB 121 GlnThrPheValLeuAlaProGlnLysSerValAlaHisLysPheTy---ValHisAsn 139



```

OY. 500 GACTGCTTCGGCTTCCAGGAC 520
Db 140 AspilepnearglyrGlnasp 146

RESULT 6
US-09-764-864-1044
; Sequence 1044, Application US/09764864
; Patent No. US20020132753A1
; GENERAL INFORMATION:
; APPLICANT: Rosen et al.
; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
; FILE REFERENCE: PT23
; CURRENT APPLICATION NUMBER: US/09/764,864
; CURRENT FILING DATE: 2001-01-17
; Prior application data removed - consult PALM or file wrapper
; NUMBER OF SEQ ID NOS: 1792
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1044
; LENGTH: 326
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-764-864-1044

Alignment Scores:
Pred. NO.: 1.54 Length: 326
Score: 89.00 Matches: 53
Percent Similarity: 30.36% Conservative: 15
Best local Similarity: 23.66% Mismatches: 62
Query Match: 5.98% Indels: 94
Db: 10 Gaps: 11

US-09-763-902B-10 (1-830) x US-09-764-864-1044 (1-326)
OY 37 ACTTCCTGACGCCCTGGTGTCCCAAGGACAGAGAAATACCCGTGTGACCCCTCTTC 96
Db 160 ThSerleuaspIndPrometProglyngly Pro-Prop 172
OY 97 CATAGAACCAGAGATGCATCTGTGGATTTCAGACCAATGTCGATCGAGCCCTCAGAC 156
Db 172 ovalGlnProgluHisAla PheargAlaGlnleuPro-HisProAlaSerT 189
OY 157 TGTCTGAGAGCTTGTCAATGTCTACTACA- 185
Db 189 hrSerSerSerThrSerSerThrThrAlaProThrAlaThrThrSerAlaSerThr 209
OY 186 -CCACCATGGATAGC- 204
Db 209 erThrProgluProglyProProThrPalasSerGlnProThrProgluArgGly 229
OY 205 GCGTT--TGTCTGCCCGCTGTACATGGGACACAGCCACCCGTGTGCAATGGCAATGC 261
Db 229 lyaAlaGlnCysGlyScysAlaProglyLTP- GluThrPasnGlyLysp- 244
OY 262 TGTTCAGGACAGAAATCCTTAGAGAGCTTTTGAATGTTGCCCTTCAGCGATTCGA 321
Db 245 - 249
OY 322 AATCAGCGTGTAGTACCTGCCAGCCCTGTCAATGATGAAACCCACACCAAGACAGCAGGT 381
Db 249 spAlaAlaLeuProThrPro- 255
OY 382 CCTTGTTCATCTGTGATCAGTGAATTTGAGGGGACAAACAGGGAGCTTCAACA 441
Db 256 - ArgGlyGlnGlnArgGlyGlnGly 265
OY 442 GA--ACTTCATCTTGAACGCGCCAGCCCTCAACCCAGCAACAACTGTGGAA 489
Db 265 rglYlnHisThrSerThrThrSerThrProThrArgThrHisProGlnHisGlnLeuScy 285
OY 490 GATCGGACAGTACTGTTCCGCTTCACAGATGGGCGGACAGTACGG- 536
Db 285 lArgHisCysAlaGlySerThrMetCysThrArgProglyProGlyLeuHisGlyVal 305

```

```

OY 537 -----GGTGGCGAGAGCTCTCTTGGCTGATTCAGCCGCTGAGCTAGCAAAAT 585
Db 305 YRHSLIERHGLYGLULYSAPHSAPHSANGINSEIIEIHRPIROTHRASVAL---ARGC 324
OY 586 GCNAACCT 593
Db 324 YSLYSPro 326

RESULT 7
US-09-814-777A-99
Sequence 99, Application US/09814777A
Patent No. US2002012415A1
GENERAL INFORMATION:
APPLICANT: KOOPMAN, Peter Anthony
APPLICANT: MUSCAT, George Eugene Orlando
TITLE OF INVENTION: NOVEL POLYPEPTIDES AND POLYNUCLEOTIDES AND METHODS OF USING THE SAME
FILE REFERENCE: 21415-0003
CURRENT APPLICATION NUMBER: US/09/814.777A
CURRENT FILING DATE: 2001-03-23
PRIOR APPLICATION NUMBER: AU P06457
PRIOR FILING DATE: 2000-03-24
NUMBER OF SEQ ID NOS: 128
SOFTWARE: Patencin version 3.0
SEQ ID NO 99
LENGTH: 462
TYPE: PRT
ORGANISM: Human
US-09-814-777A-99

Alignment Scores:
Pred. NO.: 3.24 Length: 462
Score: 86.50 Matches: 73
Percent Similarity: 27.30% Conservative: 22
Best Local Similarity: 20.98% Mismatches: 118
Query Match: 5.81% Indels: 135
DB: 10 Gaps: 15

US-09-763-902B-10 (1-830) x US-09-814-777A-99 (1-462)
OY 39 TTCCTGAGAGCCCTGTTCCCAAGCAGC---AGGAAATACCTGTGAGCCCTC--- 92
Db 143 PHEPROALALASERGLYSERIALARGALAPHEARGLYSERPROALAGLYALAGLUPHE 162
OY 93 -----CTTCATAGACACGAGATGCGATCGTGAGATTTCAGACCC----- 134
Db 163 ASPGLYLEUGLYLEUPROTHPROGLIARGLYSERPROLEUNSPGLYLEUGLUPROGLYGLU 182
OY 135 -----ATGTGATCAGAGCCCTGAGAGCTGTGAGAGTTGTCATATGCTACTACACCA 188
Db 183 ALAALAPHEPHEPROPROALALAPROARGTHRALARGTRPATGPROSERALAPRO 202
OY 189 CCATGAGATTAACGGGGCGCTTGTCTGTCCCGCCGTACATAGGGCAGACGCCACTGTCT 248
Db 203 PROTHRALAHLSARGVALVALALAGLYPROARGTRGLEULARGGLYSERPROGLYGLY 222
OY 249 GGAATGGCAATGCTGTTTCAGGACAGACATCTGATGATGATTTTTGAATGT----- 302
Db 223 GLYALA-----GLNASPARALAPROARGALALARGSERLEUALACYSTHTR 238
OY 303 -----TGCCCTTCAGCGAGTTCACAAATCA 326
Db 239 ALAPROTRIPALARGPROALARGTHPRPROALARGCYSLARGARGARGPROARGPROARG 258
OY 327 GCGGTGAGACT-----GCCAGCCTGTTCATGATGAAGCCACAC 365
Db 259 ARGTRIPARGALAPROSERPROGLYALARGARGSERVALGLYALARGARGGLYPROHIS 278
OY 366 CAACCCAGACACGGGTCTGTTGTGTCATCTGCTGGATCAGTGAAGATTGGAGGGAACAAC 425
Db 279 ARGVALARGPRO-----VALPROGLIENGLINPROASPSERIALARGARGPROARG 295
OY 426 AACGGAGCTTCAACAGACACTTCATCTGACCGCCAGGCGCTCAC----- 470

```

Db 296 AlaProValProArgLysIleThrGlyIleThrGlyProAlaArgHisValLeuProArgGly 315  
OY 471 -----CGAGCAGACAGAGT-----GGAAGATCCGACAGTGCCTGCTCCGCTCC 515  
Db 316 GluGlnProAspSerThrSerGlnProAlaAlaArgSerIleThrAlaArgAlaSer 335  
OY 516 AGAGCT -----GGCCAGCTAGTGGGGGTGGCAGAGT----- 548  
Db 336 ArgAlaAlaAlaProProGlyProCysSerAlaSerSerArgSerProArgAspArgSer 355  
OY 549 -----CTTTGCTTCATTCAGCCCTAGC 572  
Db 356 GluProAlaValArgSerAlaLeuSerTyAlaCysMetPheGlySerMetSerGlnPro 375  
OY 573 TCTGTAGAGAAATGCACAACTCCACTCTCAAGATGAGGAAACAAAGTCAATTCTGT 632  
Db 376 ProArgSerGlnCysGlyPro----- 382  
OY 633 TGTTCGGAGACACTGCAGACTGCAGTGCAGTTCGACTTTTGTGTCAG 692  
Db 383 CysAlaArgSerThrSerGlnAlaThrLeuProGlyLeu----- 395  
OY 693 TTCTAGAGAGTCCCTTCGAAATATATACTTGTGTCA----- 731  
Db 396 -----LeuAlaThrCysLeuIleGlyProLeuArgGlyCysLeu 408  
OY 732 TACTTTCCT----- 740  
Db 409 GluPheProArgValProGlyLeuPheGlnGluAlaArgAlaGlnAspLeuAlaGlu 428  
OY 741 -----TTTCAAGTAGTAA 755  
Db 429 LeuProGlyLeuHisPheSerThrCysSerPheSerCysSerValPheSerThrThrArg 448  
OY 756 CTTTCTATTTCCTACTTGCCCA 779  
Db 449 LeuTyThrPheLeuLeuCysPro 456

RESULT 8  
US-09-763-495-2  
Sequence 2, Application US/09907495  
Patent No. US20020081696A1  
GENERAL INFORMATION:  
APPLICANT: Meyers, Rachel  
TITLE OF INVENTION: 32529, A NOVEL HUMAN GUANINE NUCLEOTIDE EXCHANGE FACTOR  
FILE REFERENCE: NMT-175  
CURRENT APPLICATION NUMBER: US/09/907,495  
CURRENT FILING DATE: 2001-07-16  
PRIOR APPLICATION NUMBER: 60/218,383  
PRIOR FILING DATE: 2000-07-14  
NUMBER OF SEQ ID NOS: 3  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 2  
LENGTH: 802  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-763-495-2

Alignment Scores:  
Pred. No.: 5.11 Length: 802  
Score: 85.50 Matches: 64  
Percent Similarity: 35.00% Conservative: 27  
Best Local Similarity: 24.62% Mismatches: 98  
Query Match: 5.74% Indels: 71  
Gaps: 14  
DB: 10

US-09-763-902b-10 (1-830) x US-09-907-495-2 (1-802)

OY 24 TCTCTCAACCTTACTTCCCGCAGCC-----CCGGTTCGCCAGGAGG 71  
Db 39 AlaLeuLysProIleProSerProValCysLeuAspLeuPheProValAlaProGluIleu 58

OY 72 AATTAACCTGGTGGAGCCCTCTTCATAGAACAGATGCATCTGCAATTTCAGA 131  
Db 59 ArgAlaProGly-----SerArgPheLeuGlyThrProAla 71  
OY 132 CCTATGGGATCAGGCCCTGCAGAGCTGCTAGAGATTTCATGCTATACACACCA 191  
Db 72 ProLeu-----GlnGlyLeuLeuThrProLeuSerProGlyLysAspThr 87  
OY 192 TGGATA-----AGCGGCGCGTT-----TGTCTCCCGCTTACATGG 230  
Db 88 GluIleThrSerGlyGlyMetArgProSerArgAlaGlySerThrProHisCysProGly 107  
OY 231 GCACAGCA-----CCCTGTCTGGAATGSCAATGCTTTCAGAGCAAG 275  
Db 108 AlaGlnProProAlaLeuGlnGlyProThrSerProArgHisThrGlnProGln----- 125  
OY 276 AATCTTGAAGTGAATTTTGAATGTTGCTTCACAGAGTCCAAATCAGGCTGAG 335  
Db 126 -----ArgArgAlaSerHisGlySerGluLysLysSerAlaThrArg 139  
OY 336 ACTGCAGCCGCTTC-----ATG 353  
Db 140 LysMetArgValIleGlnArgGluGluValProGlyCysProGluAlaHisAlaValPhe 159  
OY 354 ATGAGCCACACCAAGCC-----AGACCAAGGCTCTTG 386  
Db 160 LeuGluProGlyGlnValAlaGlnGlnAlaLeuSerThrGlnGluProArgValGlu 179  
OY 387 TTGTATCTGTGGATCAGTGAATTTGAGGGGAGCAACAAACAGGACTTCACAGAACT 446  
Db 180 LeuSerGlySerThrArgValSerLeuGlnGlyProGluArgArgPheSer---Ala 198  
OY 447 TCATCTGACCGCCCGCCCTCACCCAGCAACAGATGTGAGATGCAAGTACTGCT 506  
Db 199 SerGluLeuMetThrArgLeuHisSerSerLeuArgGlyArgAspSer---AlaAla 217  
OY 507 TCCCTTCACAGAGTGGGCGAGTACTAGTGGGGGTGGCAGA---GGTCTTTGCTTCATTC 563  
Db 218 ArgAlaLeuIleSerGlySerGlyThrGlyAlaAlaArgGlnGlyLysAlaSerGlyMet 237  
OY 564 AGCCCTAGCTGTAGAGAAATGCACAACTGCAGCTTCAGAGATGAGAACACACTT 623  
Db 238 GluAlaArgSerValGluMetSerGlyAspArgValSerArgPro----- 252  
OY 624 CATTTCTGTGTGGCGAGACT-----GCAGACTCCACTGTCGAGGTGAACTCT 677  
Db 253 -----AlaProGlyAspSerArgGlnGlyLysPheArgLeuAspThr 269

RESULT 9  
US-09-746-491-8  
Sequence 8, Application US/09746491  
Patent No. US20020137202A1  
GENERAL INFORMATION:  
APPLICANT: Burgess, Catherine E.  
TITLE OF INVENTION: No. US20020137202A1 Proteins and Nucleic Acids Encoding Same  
FILE REFERENCE: 15966-621  
CURRENT APPLICATION NUMBER: US/09/746,491  
CURRENT FILING DATE: 2000-12-20  
PRIOR APPLICATION NUMBER: USSN 60/171,329  
PRIOR FILING DATE: 1999-12-21  
NUMBER OF SEQ ID NOS: 72  
SOFTWARE: Patent In Ver. 2.0  
SEQ ID NO 8  
LENGTH: 1604  
TYPE: PRT  
ORGANISM: Homo sapiens  
US-09-746-491-8

Alignment Scores:  
Pred. No.: 6.68 Length: 1604  
Score: 85.50 Matches: 64  
Percent Similarity: 35.00% Conservative: 27

Best Local Similarity: 24.628  
Query Match: 5.748  
DB: 10  
Gaps: 14  
Mismatches: 98  
Indels: 71

US-09-763-902B-10 (1-830) x US-09-746-491-8 (1-1604)

```

OY 24 TCCTTCAACCTTACTTCCCTCAGCC-----CCTGGTTCCTCCAGCAGAG 71
DB 841 AAlaLeuIysProPseProValYAlYsLeuAspLeuPheProValAlaProIuInduLeu 860
OY 72 AAATACCTGGTGGAGCCCTCCCTCCATAGAACACAGATGCTGATTCAGAGA 131
DB 861 ArgAlaProGly-----SerArgTyrSerLeuGlyThrProAla 873
OY 132 CCTATGTGATCAGCCTGCAGAGCTGCTGAGAGTTGTCAATGCTACTACACCACA 191
DB 874 ProLeu-----GlnGlyLeuThrProLeuSerProGlyGlySerAspThr 889
OY 192 TGGATA---AGCGGGGGGTT-----TGCTGCCCGCCCTGTACATGG 230
DB 880 GluIleThrSerGlyGlyMetArgProSerArgAlaGlySerTyrProHisGlyProGly 909
OY 231 GCACAGCCA-----CCCTGGTCTGGAATGGCAATGCTGTTTCAGACAG 275
DB 910 AlaGlnProProAlaLeuGluGlyProTyrSerProAlaGlnHisThrGlnProGln----- 927
OY 276 AATCCTGATGAGTGTTTTGAATGTTGCTCCAGAGAGTTCCAAATCAGCGGTGTAG 335
DB 928 -----ArgArgLaserHisGlySerGlyLysSerAlaThrArg 941
OY 336 ACTGCCACCTGTTC-----ATG 353
DB 942 LysMetArgValItyGlnArgGluGluValProGlyCysProGluAlaHisAlaValPhe 961
OY 354 ATGAAGCCACACACAGCC-----AGACCCAGGCTCTTG 386
DB 962 LeuGlnProGlyGlnValItyGlnGluGlnAlaLeuSerThrGlnGluProArgValGlu 981
OY 387 TTTCATCTGTGGATCAGTGAAGTTTGAGGGGAAACAAACAGGACTTCAACAGAACT 446
DB 982 LeuSerGlySerThrArgValSerLeuGluGlyProGluArgArgPheSer---Ala 1000
OY 447 TCATCCTGACGCGCCAGCCCTCACCCACAGACAGATGTGAAGATCCCAAGTACTGCT 506
DB 1001 SerGluLeuMetThrArgLeuHisSerSerLeuArgLeuGlyArgAsnSer---AlaAla 1019
OY 507 TCCCTTCACAGAGTGGCCAGTACAGTGGGGTGGCAGA---GGTCTCTTTCCTCATTC 563
DB 1020 ArgAlaLeuIleSerGlySerGlyThrGlyAlaAlaArgGluGlyLysAlaSerGlyMet 1039
OY 564 AGCCCTAGCTCTGAGAAATGCAACCTGACTCTCAGATGATGTAGAAACAGATT 623
DB 1040 GlnAlaArgSerValGlnMetSerGlyAspArgValSerArgPro----- 1054
OY 624 CATTTCTGTTGTTCCGAGACT-----GCAGACTCCACTGTCGAGGTTGAACTCT 677
DB 1055 -----AlaProGlyAspSerArgGluGlyAspTyrSerGluProArgLeuAspThr 1071

```

LENGTH: 197  
TYPE: PRT  
ORGANISM: Gallus gallus  
US-10-149-326-8

Alignment Scores:  
Pred. No.: 4.84  
Score: 83.50  
Percent Similarity: 34.78%  
Best Local Similarity: 27.95%  
Query Match: 5.618  
DB: 9  
Gaps: 8  
Length: 197  
Matches: 45  
Conservative: 11  
Mismatches: 45  
Indels: 60

US-09-763-902B-10 (1-830) x US-10-149-326-8 (1-197)

```

OY 188 ACCATGATTAAGCGGGCGGCTTGTCTGTCGCCCTGTACATGGCCACAGCCCTGTGTC 247
DB 27 ThrMetAspAlaArg----- 31
OY 248 TGCATGCGCAATGCTGTTCAGA-----CAAGATCCTGAGTGTTTT 295
DB 32 TrpLeuAspAsnValGlySerGlyAspLeuProAspAspGluAspIleGlyIuPheThr 51
OY 296 GAATGTTGCTTCCAGGAGATCCCAATC----- 325
DB 52 ProHisLeuThrSerAspGluPheAspIleAspAspThrSerGlySerGlyAspTyrSer 71
OY 326 -----AGCGTGTAGACTCCACCTGTTACATGATAGACC 361
DB 72 AspTyrAspAspAlaIleTyrLeuThrThrValAspThrProAlaIleSerAspAsnTyr 91
OY 362 ACACCAACCCAGACCCAGCTCTGTGTCTGATCAGTGAAGTTGAGGAGAAC 421
DB 92 IleProGlyAspThr-----GluArgLysMetGluGlyIu 103
OY 422 AAACACGGGACTTCAACCAAGACTTCACTGACCGCCAGGCTCACC----- 472
DB 104 LysLysAsnThrMetLeuAspAsnGluIleIleProAspLysAlaSerProValGluAla 123
OY 473 -----AGCACACAGTGTGAGATGCCAGTACTCTTCCTGCTCC----- 515
DB 124 AsnLeuSerAsnLysIle---SerMetLaser---ThrAlaAsnSerIlePheGluArg 142
OY 516 -----AGACTGGGCGCAGCTAGTGGGGTGGCAGAGT---CTCTTCTGCTACTT 562
DB 142 GThGluValLeuThrAlaLeuIleAlaGlyLysAlaValGlyLeuLeuPheAlaValPhe 162
OY 563 C 563
DB 162 e 162

```

RESULT 11  
US-09-741-669-458  
Sequence 458, Application US/09741669  
Patent No. US20020022718A1  
GENERAL INFORMATION:  
APPLICANT: Forsyth, R. Allen  
APPLICANT: Ohlsen, Karl L.  
TITLE OF INVENTION: Genes identified as required for  
FILE REFERENCE: ELITRA 009A  
CURRENT APPLICATION NUMBER: US/09/741,669  
PRIORITY FILING DATE: 2000-12-19  
PRIORITY FILING DATE: US 60/173005  
PRIORITY FILING DATE: 1999-12-23  
NUMBER OF SEQ ID NOS: 481  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 458  
LENGTH: 475  
TYPE: PRT  
ORGANISM: Escherichia coli  
US-09-741-669-458

## Alignment Scores:

Pred. No.:	6.8	Length:	475
Score:	83.50	Matches:	49
Percent Similarity:	40.20%	Conservative:	33
Best Local Similarity:	24.02%	Mismatches:	49
Query Match:	5.61%	Indels:	73
DB:	10	Gaps:	10

US-09-763-902b-10 (1-830) x US-09-741-669-458 (1-475)

```

QY 62 AAGGACAGAAATACCTGGTGGAGCCCTTCATGAGACAGAGATGGATCTGTG 121
    |||||
DB 216 Glnserglusermetprrtprtyrserclprrthrlleuaglvalleuaglthrv 235
    |||||
QY 122 GATTCAAGACCTTGTGGAT---CAGGCTGCAGAGCTGCTGAGAGATTGTCAATGTC 178
    |||||
DB 236 Glnlelglarvalvalasplaglprmetarppheprovalglntyrvalasn--- 254
    |||||
QY 179 TACTACACACCATGAGTAAGCGGGGCTTGGCTGCTCCGCTGATGAGGACACCC 238
    |||||
DB 255 -----AtrpProasnleuaspphearglytyrAlaglythr--- 266
    |||||
QY 239 ACCCTGGTGGAAATGGCAATGCTGTTCAGACAGAAATCCTGAGTGGATTTTGA 298
    |||||
DB 267 -----LeuAlaserglyargvalgluvalgllythrvalys 279
    |||||
QY 299 ATGTGGCTTC--- 310
    |||||
DB 280 ValleuprosarglyvalgluserasnvalAlaarglyvalthrPheasplyasparg 299
    |||||
QY 311 -----AGGAGTTCCAATCGC 328
    |||||
DB 300 GluGlualaphleaglgluAlaAlaethrleuvalleuthraspdluleasplleser 319
    |||||
QY 329 GTGTGAGACGCCACCTGTCATGATGAGCCACACCAAGC---CAGACCAAGCTCTT 385
    |||||
DB 330 ArgglyaspleuleuleuAlaAlasplunlaleuProAlaValGlnserAlaserVal 339
    |||||
QY 386 GTGTGATCTGTGATCAGTGA--- 409
    |||||
DB 340 Aspalval---TrrpetaAlagluginProleuserProglylnserTyrAspllelyst 359
    |||||
QY 410 TTTGGGGGAGAACAA---ACGGAGCTTCACCAAACTTCATCTGACC 457
    |||||
DB 359 leaAlaglylsylsThrargAlaargvalasplglylleargtyrGlnval-----AspI 377
    |||||
QY 458 GCCCAGGCTCACCCAGACACAGACAGTGGAGATCGCAGAGTACTGCTCCGCTTCAG 517
    |||||
DB 377 leasnasnleuthrlnarG---GluValGluasn-----LeuProleuasnG 392
    |||||
QY 518 GACTGGCC 525
    |||||
DB 392 lytlegly 394
    |||||

```

## RESULT 12

```

US-10-273-680-2
; Sequence 2, Application US/10273680
; Publication No. US20030099996A1
; GENERAL INFORMATION:
; APPLICANT: Millennium Pharmaceuticals, Inc.
; APPLICANT: Tomlinson, James E.
; TITLE OF INVENTION: VELPZ, A VASCULAR ENDOTHELIAL CELL
; TITLE OF INVENTION: SPECIFIC AND LIM DOMAIN CONTAINING MOLECULE AND USES
; FILE REFERENCE: MP101-317P1RNM
; CURRENT APPLICATION NUMBER: US/10/273,680
; CURRENT FILING DATE: 2002-10-18
; PRIOR APPLICATION NUMBER: US 60/329,756
; NUMBER OF SEQ ID NOS: 7
; SOFTWARE: FastSeq for Windows Version 4.0

```

```

; SEQ ID NO 2
; LENGTH: 831
; TYPE: PRT
; ORGANISM: Homo Sapiens
US-10-273-680-2

```

## Alignment Scores:

Pred. No.:	17.5	Length:	831
Score:	80.50	Matches:	59
Percent Similarity:	32.08%	Conservative:	26
Best Local Similarity:	22.26%	Mismatches:	87
Query Match:	5.41%	Indels:	93
DB:	9	Gaps:	13

US-09-763-902b-10 (1-830) x US-10-273-680-2 (1-831)

```

QY 83 TGGACCCCTCCTTCATGAAACAGAGATGGCATGTGGATTTCAGACCTATGTGAT 142
    |||||
DB 35 TrpValProProglyLeuArgProGlnGlnleuLeuTyrPhe----- 49
    |||||
QY 143 CAGGCTGCAGAGCTGAGGAGCTTGTCAATGCTACTACACCATGATAGCGG 202
    |||||
DB 50 ---AlacysLeuProGlnGlnlyVal-----ProTyrValasnserProglylyls 66
    |||||
QY 203 CGGCTGTGTCTCCCTGCTGATGAGGACACACCATGCTGTGGAATGGCAATGCT 262
    |||||
DB 67 HlsArglylelysglnleuLeuTyrGln-----LeuProProHlsaspsnglu 82
    |||||
QY 263 GTTTCAGACAGAAATCCTTGTGATGAGTTTGTGAATGTTCCCTCCAGCAGTTCCAA 322
    |||||
DB 83 ValArgTyrCysGlnserleuserGlu-----GluGlnlylsysglu 96
    |||||
QY 323 ATCAGCGTGTGACTGGCAGGCTGTTCATGATGAAGCCACACCAAGCAGAC--- 376
    |||||
DB 97 LeuGlnValPheSerAlaGlnArg---LysLysGlnAlaLeuGlyArgGlyThrIlelys 115
    |||||
QY 377 -----ACGTCCTTGTGTGATCTGTGATGATGAGTGGAGGAGGAGAA 424
    |||||
DB 116 LeuLeuserArgAlaValMethIshAlaValCys-GluGlnCysGlyLeuLys---Lleas 134
    |||||
QY 425 CNAAGGACTTCACACAGAACTTCATCTGACGCCGCCAGGCTCACCCAGCAACACAGTG 484
    |||||
DB 134 ngllygly----- 136
    |||||
QY 485 TCGAAGATCGCAAGTACTGCTTCGCTTCGAGACTGGGCCAGACTAGTGGGGTGCGAG 544
    |||||
DB 137 -----GluValAlaValAlaPheAlaSerArgAlaGlyProGlyValCysTrpHispr 153
    |||||
QY 545 AGGTCTCTTGTCTTCATTAGCCCTAGCTCTGTA----- 578
    |||||
DB 153 osercysPheValCysPheThrCysasnGlnleuLeuValAspleuIleTyrPheTyrG 173
    |||||
QY 579 -----GAGAAATGCMAAATCTGACTTCAG 604
    |||||
DB 173 nasplglylsIleHscysglyArgHisHlsAlaGlnleuLeuLysProArgCysSerAl 193
    |||||
QY 605 GATGTGAGGAGACAAAGT----- 623
    |||||
DB 193 acys-----AspGlnleuIlePheAlaAsplucysThrGlnAlaGlnlyArgHisTr 211
    |||||
QY 624 -----CATTTCTGTGT-----TGGGAGACACTG-----CAGACTCCAC 658
    |||||
DB 211 phlsMetlyshsPheCysCysleuGlnCys-GlnThrValleuGlyGlyGlnArgTyrI 231
    |||||
QY 659 TGTGCCAGAGTGAAGTCTTTTGTGTGCTCAAGTCTTGAAGTCCCTTCTCTGAATATA 718
    |||||
DB 231 leMetlysaspllyArgProPheCysCysglyCysPheGlnserleuTyrAlaGlnTyrC 251
    |||||
QY 719 TACTTGTGTGT 729
    |||||
DB 251 ysglnthrcys 254
    |||||

```

## RESULT 13

US-09-982-610-24  
Sequence 24, Application US/09982610  
Patent No. US20020146420A1  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc.  
Bennett, Brian D.  
Goeddel, David  
Lee, James M.  
Matthews, William  
Tsai, Siao Ping  
Wood, William I.  
TITLE OF INVENTION: PROTEIN TYROSINE KINASE AGONIST ANTIBODIES  
NUMBER OF SEQUENCES: 45  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Winpatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/982,610  
FILING DATE: 17-Oct-2001  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/446,648  
FILING DATE: 1996-MAY-23  
APPLICATION NUMBER: 08/222616  
FILING DATE: 04-APR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Lee, Wendy M.  
REGISTRATION NUMBER: 40,378  
REFERENCE/DOCKET NUMBER: P0821P3PCT  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-1994  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 24:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1276 amino acids  
TYPE: Amino Acid  
TOPOLOGY: Linear  
SEQUENCE DESCRIPTION: SEQ ID NO: 24:  
US-09-982-610-24  
Alignment Scores:  
Pred. No.: 23.4 Length: 1276  
Score: 80.00 Matches: 59  
Percent Similarity: 32.75% Conservative: 34  
Best Local Similarity: 20.77% Mismatches: 86  
Query Match: 5.37% Indels: 105  
Gaps: 14  
US-09-763-902b-10 (1-830) x US-09-982-610-24 (1-1276)  
QY 33 CCTTACTCCGTCAGACCCCTGTCCTCCAGACGAGGAAATACCT--GATGAGCC 89  
DB 985 PROGLNAPRVALARGYLTHPRPHISPRDARGASPTHLASERPHESERGLYALA 1004  
QY 90 CACCTTCATACACAGATGCGATCTGTGATTTCAAG-----CCTATGTCG 140  
DB 1005 -----GluTrpIsglyLeuThrGluAlaProSerProValProAlaGTP 1018  
QY 141 ATCAGGCTGCAGAGAGTGTGAGAGTTGTCAATGTCTACACTACACACACAGGATTAAC 200  
DB 1019 IIEALALEALALARGLYVALRYSER-----TTPGLNPE 1030  
QY 201 GGCGGCGTTTGCTGCCGCTGTACATGGGACACACACCTGTGTGAATGGCAATG 260

DB 1031 Gly-----GluThrGlyPheGlyLysSerAlaIle 1040  
QY 261 CTGTTTCAGACAGAAATCCTTGAGTGGATTTTAAATGTTGCTTCAGGAGATTCC 320  
DB 1041 IIEGLYGLYLUNSHISPRO-----ProAlaThrSer 1051  
QY 321 AAATCAGCGTGTAGACTCCAGCCCTGTTCATGATGAGACACACACACACCA--- 377  
DB 1052 GLYASNSER-----ArgProAlaGlyValArgAlaProPhe 1062  
QY 378 -----CGGTCTTGTCTCTCTGTGATGATGAGATTGAGG 416  
DB 1063 PROGLNAPRPRVALPROGLNGLYSGLYVALPROASRILESERLIN----- 1078  
QY 417 GGAACAACACAGGAGCTTCAACACACACTATCCTGACCGCCAGG-----CCTCAC 470  
DB 1079 -----ProGlnValProProSer 1085  
QY 471 CCAGCACACAGTGTGGAGAA--TCGCAAGTACTGCTCCGCTTCAGAGACTGGCCA 527  
DB 1086 PTPRPRVALARGSERARGPRGOLYSARGVALLEUPROCYSGLNLEUINSERGLYGLY 1105  
QY 528 GCTAGTGGGGGTGGCAGAGGCTCTTCTGCTTCAATGACCCCTAGCTCTAGAGAAATGC 587  
DB 1106 ALAVALPROGLYGLYLS-----LysGlyCys 1114  
QY 588 AAACCTCGACTCTCAAGATGTGAGGACACAAAGTTATTCGTGTGGCGAGAGACT 647  
DB 1115 GINGLYPROVALTHRYSERLEUGLYPHEVALVALPROTHRCYSCYSHISHISGLN 1134  
QY 648 GCAGACTCCACTGTGCCGAGGTGAACCT----- 677  
DB 1135 THRINSERPHESERLEUVALASNAIAPROPROALALALAPHEILEUYS 1154  
QY 678 -----TTTTGTGCTCAAGTCTTAGAGACTCCCTTCTGAAATATATCTGTT-- 726  
DB 1155 VALPHEGLUPHE-CYPHERTPSERPHEPHEPRORPHEPCYSPHEPHEVALLE 1174  
QY 727 -----TGCATAGTTTCCTTTTCAAGAGTAGTAATTTTCTATTCTTCT 770  
DB 1174 UPHERPETHYRARGPROCYSHISASNPHEVALLEUGLYTHRCPYSPHERMETALASE 1194  
QY 771 ACTTGCCACAG 780  
DB 1194 RPHEALAGLN 1197  
RESULT 14  
US-10-160-501-17  
Sequence 17, Application US/10160501  
Publication No. US20030059919A1  
GENERAL INFORMATION:  
APPLICANT: Meyers, Rachel E.  
APPLICANT: Rudolph-Owen, Laura A.  
APPLICANT: Kapeller-Lieberman, Rosana  
TITLE OF INVENTION: NOVEL HUMAN 39228, 21956, 25856, 22244, 8701, 32263,  
TITLE OF INVENTION: 50250, 55158, 47765, 62088, 50566, AND 48118  
TITLE OF INVENTION: MOLECULES AND USES THEREFOR  
FILE REFERENCE: MNT-250  
CURRENT APPLICATION NUMBER: US/10/160,501  
CURRENT FILING DATE: 2002-05-30  
PRIOR APPLICATION NUMBER: US 09/838,573  
PRIOR FILING DATE: 2002-04-18  
PRIOR APPLICATION NUMBER: US 60/197,747  
PRIOR FILING DATE: 2000-04-18  
PRIOR APPLICATION NUMBER: US 09/870,133  
PRIOR FILING DATE: 2001-05-29  
PRIOR APPLICATION NUMBER: US 60/207,649  
PRIOR FILING DATE: 2000-05-26  
PRIOR APPLICATION NUMBER: US 09/870,130  
PRIOR FILING DATE: 2001-05-29  
PRIOR APPLICATION NUMBER: US 60/207,640  
PRIOR FILING DATE: 2000-05-26





GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 24, 2003, 19:26:18 ; Search time 27.5127 Seconds

(without alignments)  
5800.355 Million cell updates/sec

Title: US-09-763-902b-10

Perfect score: 1488  
Sequence: 1 tgaacttgacacacgctg.....ataatttaataacacacg 830

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 566448

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ n2p.model -DEV=xip  
-O/cgnt2.1/USPTO.spool/US09763902/rnat.24062003.130341.12467/app\_query.fasta.1.2254  
-DB=PIR\_73 -OEMT=fastan -SUFFIX=n2p.rpr -MINMATCH=0.1 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsu62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USRR=US09763902.ccg.1.1.108.grnat.24062003.130341.12467 -NCPU=6 -ICPU=3  
-NO\_MMP -IARECOVERY -NEG\_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6  
-Fgapext=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

PIR\_73:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	124	8.3	127	2	S00751 placental protein
2	118	7.9	522	2	protein T23J18.22
3	117.5	7.9	126	2	probable nuclear t
4	116.5	7.8	125	2	nuclear transport
5	107	7.2	123	2	probable nuclear t
6	106	7.1	537	2	hypothetical prote
7	105	7.1	122	2	protein F17I21.10
8	103.5	7.0	133	2	hypothetical prote
9	91.5	6.1	133	2	probable RNA-bindi
10	90.5	6.1	434	1	suppressor 2 prote
11	88.5	5.9	353	2	hypothetical prote
12	87	5.8	485	2	hypothetical prote
13	86.5	5.8	1309	2	probable RNA-dirc
14	86	5.8	328	2	glycerol 3-phospha

C 15	86	5.8	7962	2	I38346
C 16	85.5	5.7	915	2	S54485
C 17	83.5	5.6	197	2	A53126
C 18	83.5	5.6	355	2	JN0795
C 19	83.5	5.6	475	1	JN0327
C 20	83.5	5.6	475	2	E91079
C 21	83.5	5.6	475	2	F85924
C 22	83	5.6	427	2	E96716
C 23	82.5	5.5	894	2	T27007
C 24	81.5	5.5	458	2	A40392
C 25	81.5	5.5	573	2	A33533
C 26	81	5.4	309	2	T25800
C 27	81	5.4	1198	2	T28678
C 28	80.5	5.4	152	2	C72653
C 29	80.5	5.4	319	2	C72729
C 30	80.5	5.4	423	2	A48152
C 31	80.5	5.4	428	2	F86270
C 32	80.5	5.4	773	2	A82998
C 33	80	5.4	143	2	E72699
C 34	80	5.4	555	1	RGASMA
C 35	79.5	5.3	135	2	G84212
C 36	79.5	5.3	185	2	T15623
C 37	79	5.3	315	2	S63382
C 38	78.5	5.3	364	2	D82424
C 39	78.5	5.3	455	2	B82525
C 40	78.5	5.3	546	2	F70866
C 41	78.5	5.3	622	2	T24632
C 42	78.5	5.3	649	1	H64476
C 43	78	5.2	296	2	UC7268
C 44	78	5.2	316	1	S27981
C 45	78	5.2	469	2	A57531

# ALIGNMENTS

elastic titin - hu  
CES1 protein - yea  
syndecan-4 precurs  
butyrate kinase (E  
sulfate adenylyl tr  
hypothetical prote  
hypothetical prote  
probable RNA-bindi  
alpha-2-adrenergic  
cell surface glyco  
C2H2-type zinc fin  
polyketide synthas  
hypothetical prote  
hypothetical prote  
zinc finger protei  
hypothetical prote  
probable oxidoredu  
hypothetical prote  
regulatory protein  
hypothetical prote  
hypothetical prote  
luxp protein VCA07  
conserved hypochet  
probable maltase -  
hypothetical prote  
endopeptidase La h  
paired-type homeod  
homoserine kinase  
EGR alpha transcr

RESULT 1	S00751	placental protein 15 - human
N/Alternate names:	PL15	
C/Species:	Homo sapiens (man)	
C/Date:	31-Dec-1988	#sequence-revision 31-Dec-1988 #text-change 24-Sep-1999
C/Accession:	S00751	
R/Grundmann, U.;	Neirlich, C.;	Rein, T.; Lottspeich, F.; Kuepper, H.A.
Nucleic Acids Res.	16	4721, 1988
A>Title:	Isolation of cDNA coding for the placental protein 15 (PP15).	
A/Reference number:	S00751; MUID:88247772; PMID:3380696	
A/Accession:	S00751	
A/Molecule type:	mRNA	
A/Cross-references:	EMBL, X07315; NID:g35578; PIDN:CAA30278.1; PID:g35579	
C/Superfamily:	Yeast nuclear transport factor NTF2	
Alignment Scores:		
Pred. No.:	0.000455	Length: 127
Score:	124.00	Matches: 31
Percent Similarity:	43.59%	Conservative: 20
Best Local Similarity:	26.50%	Mismatches: 54
Query Match:	8.33%	Indels: 12
DB:	2	Gaps: 3
US-09-763-902b-10 (1-830) x S00751 (1-127)		
QY	167	TTTGTCATGCTGTACTACACCAACATGGATGACGGCGGTTGCTGTCGCCGCTGAC 226
DB	14	PheileglnhstyttyglInleuPhespsnasparqthngInleuGlyAlaIlelyr 33
QY	227	ATGGCAGACGACCCCTGCTGTCGATGCAATGCTGTTTCAGACGAAATCTTGAGT 286
DB	34	IleasparlaserCysleuThrtrpInuGlnInphelngInlyLeuAlaIleIval 53
QY	287	GAGTTTGTGAATGTCCTTCAGCAGATTCACAAATCAGCGTGTA-----GACTGC 340
DB	54	GluIysleuSerSerleuProphelInlyslleGlnhstSerlleThrAlaGlnsPhls 73



```

OY 341 CAGCCTGTCATGATGAGAACCCACACACAGCAGCAGCTGCTGTCATCTGCA 400
    |||||
DB 74 GlnProThrProAsp-----SerCysIleIleSerMetValValGly 87
OY 401 TCAGTGAAGTTTGGGGGACAAACACGCGGACTTCACACAGACTTCCTGACCGCC 460
    |||||
DB 88 GlnLeuLysAlaAspGlnAspProIleMetGlyPheHisGlnMetPheLeu----- 105
OY 461 CAGGCTCACCACGACACAGACTGCGAAGATCGCAAGTGGCTTCCG 511
    |||||
DB 106 -----LysAsnIleAsnAspAlaTrpValCysThrAsnAspMetPheArg 120
RESULT 2
H86248
Protein T23J18.22 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: H86248
R:Thellogis: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,
ansen, N.F.; Hughes, B.; Hulzar, L.
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matli, R.; Marzella,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712.
A:Accession: H86248
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-522 <STO>
A:Cross-references: GB:AE005172; NID:96554189; PIDN:AAF1635.1; GSPDB:GN00141
C:Genetics:
A:Gene: T23J18.22
A:Map position: 1
Alignment Scores:
Pred. No.: 0.00224 Length: 522
Score: 118.00 Matches: 38
Percent Similarity: 43.448 Conservative: 15
Best Local Similarity: 31.158 Mismatches: 49
Query Match: 7.928 Indels: 20
Gaps: 4
US-09-763-902B-10 (1-830) x H86248 (1-522)
OY 98 AATGACACGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 157
    |||||
DB 335 TTTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 342
OY 158 GCTGAGAGATTGTCATATGTCATACACACCATGATAGCGCGGCTTGTGCTCC 217
    |||||
DB 343 -----PheValAsnHisTyrTrpHisLeuPheAspAsnAspArgSerSerLeuSer 359
OY 218 CGCTGTACATGGGACGACGACCGCTGCTGATGATGATGATGATGATGATGATGAT 277
    |||||
DB 360 SerLeuTyrAsnProThrSerLeuLeuThrPheGlnGlyGlnThrIleTyrGlyValAsp 379
OY 278 TCCTGATGATGATTTTGAATGTTGCTTCACGAGTTCCAA-----ATCAGCGCTG 331
    |||||
DB 380 AsnIleSerAsnLysLeuLysGlnLeuProPheAspGlnCysHisHisLeuIleSerThr 399
OY 332 GTAGACTCGACGCTGTCATGATGATGATGATGATGATGATGATGATGATGATGAT 391
    |||||
DB 400 ValAspSerGlnPro-----SerSerMetAlaGlyGlyGlyGlyIleLeuValPhe 417
OY 392 ATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 448
    |||||
DB 418 ValSerGlySerIleGlnLeuHisGlyGlnAspHisProLeuAlaGlnPheSerGlnThrPhe 437

```

```

OY 449 ATCTGTG 454
    |||
DB 438 HisLeu 439
RESULT 3
B86405
Probable nuclear transport factor 2 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: B86405
R:Thellogis: A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon,
ansen, N.F.; Hughes, B.; Hulzar, L.
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar,
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Matli, R.; Marzella,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712.
A:Accession: B86405
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-126 <STO>
A:Cross-references: GB:AE005172; NID:911024877; PIDN:AAG26961.1; GSPDB:GN00141
C:Genetics:
A:Map position: 1
A:Superfamily: yeast nuclear transport factor NTF2
Alignment Scores:
Pred. No.: 0.00207 Length: 126
Score: 117.50 Matches: 37
Percent Similarity: 47.978 Conservative: 22
Best Local Similarity: 30.088 Mismatches: 49
Query Match: 7.898 Indels: 15
Gaps: 5
US-09-763-902B-10 (1-830) x B86405 (1-126)
OY 158 GCTGAGAGATTGTCATATGTCATACACACCATGATAGCGCGGCTTGTGCTCC 217
    |||||
DB 10 SerLysAlaPheValAluHisTyrTrpHisLeuPheAspAsnAspArgValGlyLeuAla 29
OY 218 CGCTGTACATGGGACGACGACCGCTGCTGATGATGATGATGATGATGATGATGAT 277
    |||||
DB 30 GlyLeuTyrGlnGlnLysSerMetLeuThrPheGlnGlyGlnIleGlnGlyValGln 49
OY 278 TCCTGATGATGATTTTGAATGTTGCT-----TCCAGGAGATGCCAATACACGCTG 331
    |||||
DB 50 SerIleValAlaLysLeuThrSerLeuProPheGlnGlnCysLysHisIleSerThr 69
OY 332 GTAGACTCGACGCTGTCATGATGATGATGATGATGATGATGATGATGATGATGAT 391
    |||||
DB 70 ValAspCysGlnPro-----SerGlyProAlaSerGlyMetLeuValPhe 84
OY 392 ATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 448
    |||||
DB 85 ValSerGlyAsnLeuGlnLeuAlaGlyGlnHisAlaLeuLysPheSerGlnMetPhe 104
OY 449 ATCTGTG-----ACCGCGGACGCTCACCACGACACAGATGTGGAATGCAAGTAC 502
    |||||
DB 105 HisLeuMetProThrProGlnGlnGlySer-----PheTyrValPheAsnAsp 119
OY 503 TGTCTCCG 511
    |||||
DB 120 IlePheArg 122
RESULT 4
S50467
nuclear transport factor NTF2 - yeast (Saccharomyces cerevisiae)
N:Alternate names: protein YER009w
C:Species: Saccharomyces cerevisiae

```

```

C:/Date: 28-Jan-1995 #sequence_revision 12-May-1995 #text_change 24-Sep-1999
C:/Accession: S50467; S41793; S72237
R:/Dietrich, F.S.
Submitted to the EMBL Data Library, December 1994
A:/Description: The sequence of S. cerevisiae cosmid 9537, 9581, 9495, 9867, and lambda
A:/Reference number: S50459
A:/Accession: S50467
A:/Molecule type: DNA
A:/Residues: 1-125 <DIE>
A:/Cross-references: EMBL:U18778; NID:9603592; PIDN:AAB64542.1; PID:9603601; MIPS:YER009W
R:/Haarer, B.K.; Petzold, A.S.; Brown, S.S.
Submitted to the EMBL Data Library, July 1993
A:/Description: Identification of mutations that are synthetically lethal with altered ye
A:/Reference number: S41793
A:/Accession: S41793
A:/Molecule type: DNA
A:/Residues: 1-84 <HAA>
A:/Cross-references: EMBL:L22204; NID:9347714; PIDN:AAB49379.1; PID:9347715
R:/Haarer, B.K.; Cordub, A.; Kweon, Y.; Petzold, A.S.; Silver, P.; Brown, S.S.
Genetics 144, 495-510, 1996
A:/Title: SC3 mutations are synthetically lethal with profilin mutations and cause defe
A:/Reference number: S72237; MUID:9704444; PMID:8889515
A:/Accession: S72237
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-84 <HAA>
A:/Cross-references: EMBL:L22204; NID:9347714; PIDN:AAB49379.1; PID:9347715
C:/Genetics:
A:/Gene: SGP:NTF2
A:/Cross-references: SGD:S0000811; MIPS:YER009W
A:/Map position: 5R
C:/Superfamily: yeast nuclear transport factor NTF2

Alignment Scores:
Pred. No.: 0.00261 Length: 125
Score: 116.50 Matches: 42
Percent Similarity: 44.85% Conservative: 19
Best Local Similarity: 30.88% Mismatches: 54
Query Match: 7.82% Indels: 21
DB: 2 Gaps: 7

US-05-763-902B-10 (1-830) x S50467 (1-125):
QY 116 TCTGTGATTCAGACACCTATGTGATCGAGCGCTGCAGACTGCTGAGAGATTGTCAAT 175
   |||:::|||||::: |||
Db 2 SerLeuAspPheAsnThr-----LeuAlaGlnAsnPheThrGln 14
   |||
QY 176 GTCTACTACACCAATGATGATAGCGGGCGGCTTGCTGCCCGCTGCATGACATGGGACCA 235
   ||||| ||| ||| ||| |||||
Db 15 PheYlTYrAsnGlnPheAspThrAspArgSerGlnLeuGlnAsnLeuYlArgAsnGln 34
   ||||| ||| ||| ||| ||| |||||
QY 236 GCCACCGTGTGCTGGAATGCAATGCTGTTCAGAGACAAGAATCCCTGAGTAGATTTT 295
   :::: ||| :::: ||| :::: ||| :::: |||
Db 35 SerMetLeuThrPheGlnThrSerGlnLeuGlnAlaYlAspIleValGlnIleu 54
   ||||| :::: ||| ||||| :::: ||| |||||
QY 296 GAATGTTGCCCTTCAGGAGTTCGAA-----ATCAGCGTGTAGACTGCCAGCTGTT 349
   ||||| :::: ||| ||||| :::: ||| |||||
Db 55 ValSerLeuProPheGlnIleValGlnIleAspArgIleThrLeuAspIleGlnPro 73
   ||||| :::: ||| ||||| :::: ||| |||||
QY 350 CATGATGAAGCCACACCAAGCAGACGAGCTGCTGTGTGTATCTGTGAGATCA----- 403
   |||:::|||||::: |||
Db 74 -----AlaSerProAsn---GlyAspValLeuValMetIleThrGlyAspLeu 89
   |||:::|||||::: |||
QY 404 GTGAGATTGAGGGAGAACAAACAGGGAGCTTCAACACGAATCTATCTGACCGGCCAG 463
   :::: ||| ||| ||||| ||||| :::: ||| |||||
Db 90 IleAspGlnGlnGlnAsnProGlnIle---PheSerGlnValPheHisLeuIle----- 106
   ||| ||| :::: ||| ||||| |||||
QY 464 GCCTACCCAGCACACAGTAGTGAAGATCGAAGTAGATGCTGCTGCCG 511
   ||| ||| :::: ||| ||||| |||||
Db 107 -----ProAspGlyAsnSerYlTYrValPheAsnAspIlePheArg 120
   ||| ||| :::: ||| ||||| |||||

```

probable nuclear transport factor 2 - fission yeast (*Schizosaccharomyces pombe*)  
C:Species: Schizosaccharomyces pombe  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 15-Sep-2000  
C:Accession: T38039; t37728  
R:McDonagall, R.C.; Rajadream, M.A.; Barrell, B.G.; Whitehead, S.; Churcher, C.M.  
Submitted to the EMBL Data Library, August 1999  
A:Reference number: T38039  
A:Accession: T38039  
A>Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-123 <MCD>  
A:Cross-references: EMBL:AL109951; PIDN:CA853052.1; GSPDB:GN00066; SPDB:SPAC1B9.01c  
A:Experimental source: strain 972n-; cosmid c189  
R:Devlin, K.; Churcher, C.M.; Barrell, B.G.; Rajadream, M.A.; Walsh, S.V.  
Submitted to the EMBL Data Library, December 1995  
A:Reference number: Z21740  
A:Accession: T37728  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 'SL', 77-123 <DEV>  
A:Cross-references: EMBL:Z68197; PIDN:CAA92380.1; GSPDB:GN00066; SPDB:SPAC1Sf9.03c  
C:Gene: SPDB:SPAC1B9.01c  
C:Genetics:  
A:Map position: 1  
A:Insertions: 3/1: 30/3; 54/3; 88/3  
C:Superfamily: yeast nuclear transport factor NTF2

Alignment Scores:

Pred. NO.:	0.0238	Length:	123.
Score:	107.00	Matches:	35
Percent Similarity:	46.28%	Conservative:	21
Best Local Similarity:	28.93%	Mismatches:	53
Query Match:	7.19%	Indels:	12
DB:	2	Gaps:	4

US-09-763-902B-10 (1-830) x T38039 (1-123)

OY	158	GCTGAGCAGNTTGTTCATATCTCTACTACACCACCATGATAAGCGCGCCCTTGCTGCC	217
Dd <td>   </td> <td>::::   </td> <td>     </td>		::::	
OY <td>8<td>AAlathrcglnphetrhglnphetyrtyglntThPh easpsersparSerglnleuSer<td>27</td></td></td>	8 <td>AAlathrcglnphetrhglnphetyrtyglntThPh easpsersparSerglnleuSer<td>27</td></td>	AAlathrcglnphetrhglnphetyrtyglntThPh easpsersparSerglnleuSer <td>27</td>	27
Dd <td>     </td> <td>::::   </td> <td>     </td>		::::	
OY <td>218<td>CGCCTGTACATGGGCACAGCACCGCTGCTGGAAATGCTGTCATGCTGTTCAGACAAGA<td>277</td></td></td>	218 <td>CGCCTGTACATGGGCACAGCACCGCTGCTGGAAATGCTGTCATGCTGTTCAGACAAGA<td>277</td></td>	CGCCTGTACATGGGCACAGCACCGCTGCTGGAAATGCTGTCATGCTGTTCAGACAAGA <td>277</td>	277
Dd <td>     </td> <td>::::   </td> <td>     </td>		::::	
OY <td>28<td>SerleutyrrarglglugluserMerleuSerPhegluglyalaglInleunglnglyThrlys<td>47</td></td></td>	28 <td>SerleutyrrarglglugluserMerleuSerPhegluglyalaglInleunglnglyThrlys<td>47</td></td>	SerleutyrrarglglugluserMerleuSerPhegluglyalaglInleunglnglyThrlys <td>47</td>	47
Dd <td>     </td> <td>::::   </td> <td>     </td>		::::	
OY <td>278<td>TCTGTGAGTGAATTGTTTGAATGCTCCCTCCAGCAGTCCAA-----ATCAGCGTG<td>331</td></td></td>	278 <td>TCTGTGAGTGAATTGTTTGAATGCTCCCTCCAGCAGTCCAA-----ATCAGCGTG<td>331</td></td>	TCTGTGAGTGAATTGTTTGAATGCTCCCTCCAGCAGTCCAA-----ATCAGCGTG <td>331</td>	331
Dd <td>     </td> <td>::::   </td> <td>     </td>		::::	
OY <td>48<td>AlalleValglulysleuValserleuProheglInrvalglInHlsArGlIeserThr<td>67</td></td></td>	48 <td>AlalleValglulysleuValserleuProheglInrvalglInHlsArGlIeserThr<td>67</td></td>	AlalleValglulysleuValserleuProheglInrvalglInHlsArGlIeserThr <td>67</td>	67
Dd <td>     </td> <td>::::   </td> <td>     </td>		::::	
OY <td>332<td>GTAAGCTCCAGACCTGTCATGATGAAGCACACCAAGCCAGACCGAGCTCTGTGTC<td>391</td></td></td>	332 <td>GTAAGCTCCAGACCTGTCATGATGAAGCACACCAAGCCAGACCGAGCTCTGTGTC<td>391</td></td>	GTAAGCTCCAGACCTGTCATGATGAAGCACACCAAGCCAGACCGAGCTCTGTGTC <td>391</td>	391
Dd <td>     </td> <td>::::   </td> <td>     </td>		::::	
OY <td>68<td>LeuaspalaglInpuo-----ThrglyThrTrglSerVallleValMet<td>82</td></td></td>	68 <td>LeuaspalaglInpuo-----ThrglyThrTrglSerVallleValMet<td>82</td></td>	LeuaspalaglInpuo-----ThrglyThrTrglSerVallleValMet <td>82</td>	82
Dd <td>     </td> <td>::::   </td> <td>     </td>		::::	
OY <td>332<td>ATCTGTGATGATGAGAACGTT---GAGGGGAACAACAAGCGAGCTTCACACGAACTTC<td>448</td></td></td>	332 <td>ATCTGTGATGATGAGAACGTT---GAGGGGAACAACAAGCGAGCTTCACACGAACTTC<td>448</td></td>	ATCTGTGATGATGAGAACGTT---GAGGGGAACAACAAGCGAGCTTCACACGAACTTC <td>448</td>	448
Dd <td>     </td> <td>::::   </td> <td>     </td>		::::	
OY <td>83<td>VAlThrglgluleuLeuLeuaspJuglInmetaIagInarGTYrSerglnValphe<td>102</td></td></td>	83 <td>VAlThrglgluleuLeuLeuaspJuglInmetaIagInarGTYrSerglnValphe<td>102</td></td>	VAlThrglgluleuLeuLeuaspJuglInmetaIagInarGTYrSerglnValphe <td>102</td>	102
Dd <td>     </td> <td>::::   </td> <td>     </td>		::::	
OY <td>449<td>ATCCGACAGCCCCAGCGCTCACCCAGCAACACACATGAGGAAGATGCGAAGGACTGCTTC<td>508</td></td></td>	449 <td>ATCCGACAGCCCCAGCGCTCACCCAGCAACACACATGAGGAAGATGCGAAGGACTGCTTC<td>508</td></td>	ATCCGACAGCCCCAGCGCTCACCCAGCAACACACATGAGGAAGATGCGAAGGACTGCTTC <td>508</td>	508
Dd <td>     </td> <td>::::   </td> <td>     </td>		::::	
OY <td>103<td>HlsleuVal-----AsnasnglsnGlyasnTYrValleuAsnspLeuPhe<td>118</td></td></td>	103 <td>HlsleuVal-----AsnasnglsnGlyasnTYrValleuAsnspLeuPhe<td>118</td></td>	HlsleuVal-----AsnasnglsnGlyasnTYrValleuAsnspLeuPhe <td>118</td>	118
Dd <td>     </td> <td>::::   </td> <td>     </td>		::::	
OY <td>509<td>CGC 511<td></td></td></td>	509 <td>CGC 511<td></td></td>	CGC 511 <td></td>	
Dd <td>     </td> <td></td> <td></td>			
OY <td>119<td>Arg 119<td></td></td></td>	119 <td>Arg 119<td></td></td>	Arg 119 <td></td>	
Dd <td>     </td> <td></td> <td></td>			

RESULT 6  
T23479  
hypothetical protein K08F4.2 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C:Accession: T23479  
R:Hembry, C.  
Submitted to the EMBL Data Library, January 1996

A:Reference number: 219746

A:Accession: T23479

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-537 <WII>

A:Cross-references: EMBL:268879; PIDN:CAA93082.1; GSPDB:GN00022; CESP:K08F4.2

A:Experimental source: clone K08F4

C:Genetics:

A:Gene: CESP:K08F4.2

A:Map position: 4

A:Introns: 66/3; 99/1; 140/3; 161/3; 338/1; 419/1

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
0.0369	106.00	537	30	31	48	18	5
Percent Similarity:	48.03%						
Best Local Similarity:	23.62%						
Query Match:	7.12%						

US-09-763-902b-10 (1-830) x T23479 (1-537)

```

155 GCTGCGAGGAGG-----TTTGCATGCTCTACTACACACCATGATTAAGCG 202
    |||||
    96 AAlaAlaGlnValAlaGlnValAlaPheCysHisGlnPheTyrIleThrValSerGlnAsn 115
    |||||
    203 CGGCGTTGCTGCTCCGCGCGCTGATGAGGACGACGCCCGCTGCTGGAATGGCAATGCT 262
    |||||
    116 ArgGlyAlaIleIleThrValPheTyrGlnHisGlnSerLysPheTyrLeuAspArgGlnVal 135
    |||||
    263 GTTTCGAGACAGATCTGATGAGTGTGTTGAAATGTTGCT---TCCAGCGAGTTC 319
    |||||
    136 ValThrGlySerGlnIleValLeuValLysLeuTyrAsnHisLeuProGluThrThrHisPhe 155
    |||||
    320 CAATCAGCGGTGTGATGACGCCACGCTGTTGATGATGAGGACACACAGGACGACGACG 379
    |||||
    156 LysIleGlnSerIleLysGlyTyrPro-----ThrProHisLysGlnGly 170
    |||||
    380 GTTCCTGTTGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATGTCATG 439
    |||||
    171 ValIleIleAsnValIleIleGlyThrValAsnLeu-----ArgProPheLeu 185
    |||||
    440 CAGAACTTCATCCTGACGCCGCCGCTGACCCACACACAGTGTGGAAGATCCCAAGT 499
    |||||
    186 GlnSerPheLeuLeuGlyGlnGlnIleGlnLysLys-----TyrTyrValGlnSer 202
    |||||
    500 GACTCTCTCCGCTTCCAGGAC 520
    |||||
    203 AspAlaPheGlnTyrIleAsp 209
  
```

# RESULT 7

H86398

protein F17L21.10 [imported] - Arabidopsis thaliana

C:Species: Arabidopsis thaliana (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 31-Mar-2001

C:Accession: H86398

R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,

Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;

Nature 408, 816-820, 2000

A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.

C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marzalli,

Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shin, P.; Southwick, A.M.; Sun, H.; Tallon,

Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.

A:Reference number: A86411; MUID:21016719; PMID:11130712

A:Accession: H86398

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-122 <STNO>

A:Cross-references: GB:AE005172; NID:99802547; PIDN:AAF99749.1; GSPDB:GN00141

C:Genetics:

A:Gene: F17L21.10

A:Map position: 1

C:Superfamily: yeast nuclear transport factor NTF2

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
0.0379	105.00	122	33	33	44	12	4
Percent Similarity:	45.45%						
Best Local Similarity:	27.27%						
Query Match:	7.05%						

US-09-763-902b-10 (1-830) x H86398 (1-122)

```

158 GCTGAGAGTTTGCATGCTCTACTACACCATGATTAAGCGCGCTTCTGTC 217
    |||||
    7 AAlaValAlaPheValGlnHisTyrTyrSerThrPheAspAlaAsnArgProGlyLeuVal 26
    |||||
    218 CGCCTGATGATGGGACACGACGCCCGCTGCTGGAATGCAATGCTTTCAGGACAGAA 277
    |||||
    27 SerLeuTyrGlnGlnGlnGlnSerMetLeuThrPheGlnGlnLysIleGlnGlnSerGln 46
    |||||
    278 TCCTTGAAGTGAATTTTGAATGTTGCT---TCCAGCGAGTTCGAATCAGCGT 331
    |||||
    47 AsnIleValAlaLysLeuThrGlyLeuProPheGlnGlnLysLysAsnIleThrThr 66
    |||||
    332 GTAGAGTCCAGCGCTTCTTCTGATGAGGACCAACAGGACGACGACGCTTGTGTC 391
    |||||
    67 ValAspCysGlnPro-----SerGlyProAlaGlyLeuLeuValPhe 81
    |||||
    392 ATCTGTGATCAGTGAATTTGAGGAGCAACAAACAGG---GACTTCAACGAACTTC 448
    |||||
    82 ValSerGlyAsnLeuGlnIleValGlnGlnHisLeuLysSerGlnMetPhe 101
    |||||
    449 ATCTGACCGCCGACGCGCTGACCCACACAGTGTGGAATGTCGCAATGTCGCTTC 508
    |||||
    102 HisLeuIleSerAsnGlnGlnLysn-----TyrTyrValPheAsnAspIlePhe 117
    |||||
    509 GCG 511
    |||||
    118 Arg 118
  
```

# RESULT 8

T23921

hypothetical protein R05D11.3 - Caenorhabditis elegans

C:Species: Caenorhabditis elegans

C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 21-Jan-2000

C:Accession: T23921

R:White, S.

submitted to the EMBL Data Library, June 1996

A:Reference number: T23921

A:Accession: T23921

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-133 <WII>

A:Cross-references: EMBL:275546; PIDN:CAA99890.1; GSPDB:GN00019; CESP:R05D11.3

A:Experimental source: clone R05D11

C:Genetics:

A:Gene: CESP:R05D11.3

A:Map position: 1

A:Introns: 33/1; 63/3; 96/3

C:Superfamily: yeast nuclear transport factor NTF2

Alignment Scores:

Pred. No.:	Score:	Length:	Matches:	Conservative:	Mismatches:	Indels:	Gaps:
0.0544	103.50	133	38	38	49	31	7
Percent Similarity:	45.95%						
Best Local Similarity:	25.68%						
Query Match:	6.95%						

US-09-763-902b-10 (1-830) x T23921 (1-133)





Db 224 nProglArgValCysaspSerCysTyrGluArgLeuAspProLeuGlnCysValLeu1 244  
 QY 396 -----CAGATGACACAAAGACCGTGTC---TGGCTGGTGT 362  
 Db 244 eaasSerIleSerAsnAlaValGlnValAlaLysHisAspValAlaAspTrp----- 261  
 QY 361 GGCTTCATCATGAAAGCGTGGCAGCTGATCCAGCTGATTGGAACTCGTGAAGCA 302  
 Db 262 ----ThCysSerArgGlyTrpLeuAsnLeuPro---ValGlyLeuSerMetGlnAspG1 279  
 QY 301 CATTTCAAAA-----AACTCACTCAAGATTCCTGT-----CCTGA 266  
 Db 279 uileTyrLysAlaAlaAsnThrLeuArgGlyTyrCysGlnValAlaArgLeuAspProG1 299  
 QY 265 AACAGCATTCGCA 253  
 Db 299 uLysSerIlePro 303

## RESULT 13

T00078  
 Probable RNA-directed DNA polymerase (EC 2.7.7.49) - Chlorella vulgaris retrotransposon  
 N:Alternate names: reverse transcriptase  
 C:Species: Chlorella vulgaris  
 C:Date: 22-Jan-1999 #sequence\_revision 22-Jan-1999 #text\_change 21-Jul-2000  
 C:Accession: T00078  
 R:HisashiYama, T.; Naitoshi, Y.; Fujie, M.; Yamada, T.  
 EMBL J. 16, 3715-3723, 1997  
 A:Title: zepp, a LINE-like retrotransposon accumulated in the Chlorella telomeric region  
 A:Reference number: 214092; MUID:97361851; PMID:9218812  
 A:Accession: T00078  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-1309 <HIS>  
 A:Cross-references: EMBL:AB008896; NID:g3063386; PIDN:BAA25763.1; PID:g3073779  
 A:Experimental source: strain C-169  
 C:Genetics:  
 A:Map position: 5  
 A:Mobile element: retrotransposon zepp  
 C:Keywords: nucleotidyltransferase

## Alignment Scores:

Pred. No.: 3 93 Length: 1309  
 Score: 86.50 Matches: 71  
 Percent Similarity: 29.35% Conservative: 20  
 Best Local Similarity: 22.90% Mismatches: 99  
 Query Match: 5.81% Indels: 120  
 Gaps: 16

US-09-763-902b-10 (1-830) x T00078 (1-1309)

QY 656 GAGCTCTGCACTGTCTCCGACAAACAGAA-----ANGAAGCTGTGTCCTCAAT 606  
 Db 968 GlyThrCysGlnArgGlyGlySerThrSerGlySerTrpGlnAspSerCysSerSerPro 987  
 QY 605 CCTTGAGAGTCGAGGTTGCACTTCTCTACAGAGCTAGGCTGAA----- 561  
 Db 988 ProSerGlnArgSer-----MetTrpArgProArgSerAlaTyrGlySerThrCys 1004  
 QY 560 -----TGACCAAGAGACCTCTGACCCACC 534  
 Db 1005 MetProThrSerSerCysGlyProIleThrCysGlySerArgGlySerArgSerAlaThrThr 1024  
 QY 533 ACTAGCTGGCCAGT-----CTTGAGAGCGGAACATCACTTCGATCTCCACACTG 480  
 Db 1025 GlnSerTrpValSerHisCysSerSerThrGlySerSerArgMetGlnSerAlaAlaLeu 1044  
 QY 479 TGT-----TGCTGG 471  
 Db 1045 CysLeuGlyGlyLeuGlyGlnAlaTrpProGlyGlnAspGlnProProAlaArgCysArg 1064  
 QY 470 GTAGAGCTGGCGGCTCAGATGAAGTCTGGTTAGATCCCGTTGTTTCCCTCAA 411

Db 1065 AlaArgProGlyAlaSerGly---SerAspGly----- 1074  
 QY 410 ACTTCAGTATGCCACAGATGACACAAAGACCGTGTGCTGGTGGTTCATCAT 351  
 Db 1075 -----AlaArgAlaArgAlaHisPro---SerGlyCysLeuAlaSerSerLeu 1089  
 QY 350 GAACAGCGTGGCAGCTGTACACACCGCTGATTGGAACTCGTGAAGCAATTTCAAA 291  
 Db 1090 GluSerAlaAsn-----ProGlyAlaAlaThrGlyArgTyrValAlaArgIleSerAsp 1107  
 QY 290 ACTTACCTCAAGATTCCTTGCCTGAACAG-----CAT 258  
 Db 1108 ProSerArgArgArgValCysLeuAlaGlnProLysValAlaProThrSerGlnProHis 1127  
 QY 257 TGCCATTCACAGACAGGAGTGGCTGTGC-----CCATTCACAGGC 219  
 Db 1128 SerLysGlnHisArgSerLeuLeuCysGlnTrpThrLysAlaLeuGlnProCysThrHis 1147  
 QY 218 GGGACAGCAAGCGCGCGCTATTCATCG----- 189  
 Db 1148 SerPheLeuArgProArgSerLeuHisThrPheSerThrGlyThrSerProGlyProGly 1167  
 QY 189 ----- 189  
 Db 1168 IleHisAlaAlaHisValSerSerAlaSerThrGlnSerSerCysThrGlnSerSer 1187  
 QY 188 -----TGGTGTAGTACAGATTCGACAACTCTCAG----- 159  
 Db 1188 ThrCysArgTrpSerSerArgGlnGlySerSerGlnGlySerSerTrpGlySer 1207  
 QY 158 CAGCTGCGAGCGCTGTACACATAGAGCTTGAATGACAGATGCATCTCTGTTCTA 99  
 Db 1208 SerSerCysArgSerHisCysHisHisArg---ArgProHisLeuTyrPheLeuCysLeu 1226  
 QY 98 TGAAGAGAGCGCTCCACAGGGGTATTCCTGCTGCTGG-----GGAACCGAGCGCTGC 45  
 Db 1227 Trp-----GlySerGlnValGlnAspProGlnAlaTrpGlyLeuGlyThrArgProAla 1244  
 QY 44 AGGAGATTAAGTTGACAGAAAGCCAGC 17  
 Db 1245 HisGlnPheValVal-ArgGlnAlaThr 1253

## RESULT 14

H75251  
 glycerol-3-phosphate dehydrogenase, MND(+)-dependent - Deinococcus radiodurans (stra  
 C:Species: Deinococcus radiodurans  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 17-Mar-2000  
 C:Accession: H75251  
 R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.  
 M.; Shen, M.; Yamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.  
 S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
 A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
 A:Reference number: A75250; MUID:20036896; PMID:10567266  
 A:Accession: H75251  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-328 <WHI>  
 A:Cross-references: GB:AE002091; GB:AE000513; NID:g6460446; PIDN:AAF12158.1; PID:g64  
 A:Experimental source: strain R1  
 C:Genetics:  
 A:Gene: DR2621  
 A:Map position: 1  
 C:Superfamily: glycerol-3-phosphate dehydrogenase

## Alignment Scores:

Pred. No.: 3 64 Length: 328  
 Score: 86.00 Matches: 62  
 Percent Similarity: 39.65% Conservative: 28  
 Best Local Similarity: 27.31% Mismatches: 77  
 Query Match: 5.78% Indels: 60  
 Gaps: 11

Alignment Scores:	
Pred. No.:	5.67
Score:	86.00
Length:	79652
Matches:	62

Search completed: June 24, 2003, 19:39:13  
Job time : 38.0127 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus.n2p model

Run on: June 24, 2003, 19:24:08 ; Search time 12.6012 Seconds  
(without alignments)  
5463.815 Million cell updates/sec

Title: US-09-763-902b-10  
Perfect score: 1489  
Sequence: 1 tgaatttgatcattcacgty.....ataattaataatcacatg 830

Scoring table:  
BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 112892 seqs, 41476328 residues  
Total number of hits satisfying chosen parameters: 225784

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=framer.n2p.model -DEV=xlp  
-Q=/cg92\_1/USPTO.spool/US09763902/runtac\_24062003\_130340\_12418/app\_query.fasta.1.2254  
-DB=SwissProt.40 -OEW=fastan -SUFFIX=n2p.rsp -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blomsun62 -TRANS=human40.cdl  
-LIST=45 -DOCALLIGN=200 -THR.SCORE=pcr -THR.MAX=100 -THR.MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USBR=US09763902.ecgn.1.1.46 -tunat\_24062003\_130340\_12418 -KCPU=6 -ICPU=3  
-NO\_MMAP -LARGEQUERY -NEG.SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV.TIMEOUT=120 -WARN.TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database: SwissProt.40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	732	49.2	140	1	NXT1_HUMAN
2	726	48.8	140	1	NXT1_MOUSE
3	584	39.2	142	1	NXT2_HUMAN
4	308	20.7	133	1	NXT1_DROME
5	206.5	13.9	137	1	NXT1_CAEEL
6	124	8.3	127	1	NXT2_HUMAN
7	123	8.3	127	1	NXT2_XENLA
8	118.5	8.0	124	1	NXT2_CANAL
9	117.5	7.9	126	1	NXT2_ARATH
10	116.5	7.8	125	1	NXT2_YEAST
11	115.5	7.8	1482	1	NXT2_HUMAN
12	115.5	7.8	1482	1	NXT2_HUMAN
13	107	7.2	122	1	NXT2_ORYSA
14	107	7.2	123	1	NXT2_SCHPO
15	103.5	7.1	1466	1	NXT2_HUMAN
16	103.5	7.0	133	1	NXT2_CAEEL
17	102.5	6.9	124	1	NXT2_NEUCR
18	101.5	6.8	465	1	G3BP_MOUSE

19	91.5	6.1	434	1	G3BP_SCHPO
20	90.5	6.1	479	1	ERF2_YEAST
21	90	6.0	479	1	KLFA_MOUSE
22	87	5.8	351	1	ROM1_BOVIN
23	86	5.8	328	1	GPD4_DEIRA
24	85.5	5.7	915	1	ZDS1_YEAST
25	83.5	5.6	197	1	SDCA_CHICK
26	83.5	5.6	355	1	BURK_CLOVE
27	83.5	5.6	475	1	CYSN_ECOLI
28	81.5	5.5	573	1	C114_MOUSE
29	80.5	5.4	423	1	GFI1_RAT
30	80	5.4	555	1	WETA_EBNTI
31	79.5	5.3	515	1	DABI_RHIL0
32	79	5.3	515	1	YN8T_YEAST
33	79	5.3	1042	1	CORI_HUMAN
34	78.5	5.3	364	1	LUXP_VIBCH
35	78.5	5.3	365	1	FCGN_MOUSE
36	78.5	5.3	622	1	YRT1_CAEEL
37	78.5	5.3	649	1	LONH_MERJA
38	78.5	5.3	725	1	MCCA_HUMAN
39	78	5.2	480	1	KLFA_HUMAN
40	78	5.2	1098	1	TSU1_YEAST
41	77.5	5.2	765	1	SIM1_MOUSE
42	77.5	5.2	2738	1	PGCV_RAT
43	77	5.2	838	1	VGLH_HSV11
44	76.5	5.1	353	1	CUPS_GALME
45	76.5	5.1	458	1	A2AC_MOUSE

## ALIGNMENTS

## RESULT 1

ID NXT1\_HUMAN STANDARD: PRT: 140 AA.

AC Q9UKK6;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 15-JUN-2002 (Rel. 41, Last annotation update)  
DE NTF2-related export protein 1 (p15 protein).

CN NXT1.

OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

OX NCBI\_TaxID:9606;

RN [1]

SEQUENCE FROM N.A. AND FUNCTION.

RP MEDLINE-20036817; PubMed-10567585;  
RA Black B.E., Levesque L., Holaska J.M., Wood T.C., Paschal B.M.;

RT "Identification of an NTF2-related factor that binds Ran-GTP and  
regulates nuclear protein export."

RL Mol. Cell. Biol. 19:8616-8624(1999).

RN [2]

SEQUENCE FROM N.A.

RP MEDLINE-21638749; PubMed-11780052;  
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,

RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,

RA Bailey J., Barlow K.F., Bates K.N., Beare D.M., Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,

RA Buck D., Burrill W.D., Butler A.P., Carter C., Carter N.P.,

RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,

RA Clegg S., Copley V.E., Collier R.E., Connor R.E., Corby N.R.,

RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,

RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,

RA Griffiths D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,

RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,

RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,

RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,

RA Levasseur M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,

RA Marsh V.L., Martin S.L., McConachie L.J., McElay K., McKerrry A.A.,

RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,

RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,

RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,

RA Rice C.M., Ross M.T., Scott C.E., Sehara H.K., Shownkeen R., Sims S.,

094260	schizosacch
094263	schizosacch
089091	mus musculus
052205	bos taurus
09176	delinococcus
05011	saccharomyc
049416	gallus gall
005619	elostriidum
023845	escherichia
019467	mus musculus
007120	rattus norv
022022	emeticella
058210	rhizobium
053741	saccharomyc
095955	homo sapien
09116	vibrio chol
061559	mus musculi
010014	caenorhabdi
058812	methanococ
096193	homo sapien
013118	homo sapien
038437	saccharomyc
061045	mus musculi
098174	rattus norv
064477	herpes simp
024998	galleria me
001337	mus musculi



RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Sycamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Treacy A., Toman A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilming L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.,  
 RT "The DNA sequence and comparative analysis of human chromosome 20.",  
 RL Nature 414:865-871(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE-Lung, Placenta, and Uterus;  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP FUNCTION.  
 RA MEDLINE=20307861; PubMed=10848583;  
 RA Osareh-Nazari B., Watson C., Black B.E., Levesque L., Paschal B.M.,  
 RA Dargemont C.;  
 RT "Rangit-binding protein Nxtl facilitates nuclear export of different  
 RT classes of RNA in vitro.";  
 RL Mol. Cell. Biol. 20:4562-4571(2000).  
 RN [5]  
 RP FUNCTION.  
 RA MEDLINE=21160285; PubMed=11259602;  
 RA Guzik B.W., Levesque L., Prasad S., Bor Y.C., Black B.E.,  
 RA Paschal B.M., Rekosh D., Hammarskjold M.L.;  
 RT "Nxtl (p15) is a crucial cellular cofactor in TAP-dependent export of  
 RT intron-containing RNA in mammalian cells.";  
 RL Mol. Cell. Biol. 21:2545-2554(2001).  
 CC -1- FUNCTION: Stimulator of protein export for NES-containing  
 CC proteins. Also plays a role in the nuclear export of UI snRNA,  
 CC tRNA, and mRNA.  
 CC -1- SUBUNIT: Preferentially binds Ran-GTP. Associates with Nxf1, Nxf2,  
 CC Nxf3 and Nxf4.  
 CC -1- SUBCELLULAR LOCATION: Nuclear. Shuttles between the nucleus and  
 CC the cytoplasm.  
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC  
 DR EMBL: AF156957; AAD54942.1;  
 DR EMBL: AL096677; CAC03437.2;  
 DR EMBL: BC000759; AAH00759.1;  
 DR EMBL: BC002687; AAH02687.1;  
 DR EMBL: BC003029; AAH03029.1;  
 DR EMBL: BC003410; AAH03410.1;  
 DR Gene: HGNC:15913; NXL1.  
 DR MIM: 605811;  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF02136; NTF2.  
 DR PROSITE: PS50177; NTF2\_DOMAIN; 1.  
 KW Transport; Protein transport; Nuclear protein.  
 FT DOMAIN 16 135 NTF2  
 FT SEQUENCE 140 AA: 15847 MW: 356FA86AC3944594 CRC64;

Alignment Scores:  
 Pred. No.: 4.87e-65 Length: 140  
 Percent: 732.00 Matches: 140  
 Score: 100.00% Conservative: 0  
 Best Local Similarity: 100.00% Mismatches: 0  
 Query Match: 49.16% Indels: 0  
 DB: 1 Gaps: 0

US-09-763-902b-10 (1-830) x NXL1\_HUMAN (1-140)

OY 110 ATGCGATCTGTGATTTCAAGACCTATGTGATCAGCGCTCAGAGCTGTGAGAGATT 169  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||  
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

DB 1 MetalaserValasphelelyfThrValaspGlnalacysArgAlaIaGlnIue 20  
 OY 170 GTCAATGCTCTCTACACCATGATGATGAGGGGGCTTGCTGCTCCGCTGACATG 229  
 DB 21 ValasVallyfThrThrMetaspysArgArgLeuLeuSerArgLeuYrMet 40  
 OY 230 GGACAGCCACCCGTGCTGGAATGCAATGCTGTTTCAGAGCAAGATCCCTGAGTGA 289  
 DB 41 GlyThrAlaThrLeuValTTPasnGlyAsnAlaValSerGlyGlnSerIue 60  
 OY 290 TTTTGAATGTGCTTCCAGCAGTTCCAATCAGCTGAGTCCAGCTGCTT 349  
 DB 61 PhepHeGluMetLeuProSerSerGlnpHeGlnIleSerValValaspGlnProVal 80  
 OY 350 CATGATGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 409  
 DB 81 HisaspGlnAlaThrProSerGlnThrThrValLeuValIleCysGlySerVallys 100  
 OY 410 TTTGAGGGAGCAACAACAGGAGCTTCACCAAGACTTCAGCTGACCCAGCCTCA 469  
 DB 101 PheGlnGlyAsnLysGlnAlaArgaspPheAsnGlnAsnPhelIeThrAlaGlnAlaser 120  
 OY 470 CCCAGCAGCAGCAGTGTGGAATGCGCAGTACGCTCCGCTCCAGACTGGCCAGC 529  
 DB 121 ProSerAsnThrValTTPlysIleAlaseraspCysPheArgpHeGlnAspTTPAlaser 140

RESULT 2  
 NXL1\_MOUSE STANDARD: PRT: 140 AA.  
 AC 09Q2V9; Q9DIP7;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE NTF2-related export protein 1.  
 GN NXL1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sclerognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=20036817; PubMed=10567585;  
 RA Black B.E., Levesque L., Holaska J.M., Wood T.C., Paschal B.M.;  
 RT "Identification of an NTF2-related factor that binds Ran-GTP and  
 RT regulates nuclear protein export.";  
 RL Mol. Cell. Biol. 19:8616-8624(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Embryo;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,  
 RA Aizawa K., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,  
 RA Saito T., Okazaki Y., Gojopori T., Bono H., Kasukawa T., Saito R.,  
 RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,  
 RA Flischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,  
 RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,  
 RA Schriml L.M., Stadl F., Suzuki R., Tomita M., Wagner L., Washio T.,  
 RA Sakai K., Okido T., Furuno M., Aono H., Balarelli R., Barsh G.,  
 RA Blake J., Botfield D., Bujana N., Carninci P., de Bonaldo M.,  
 RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombereis P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyokawa K., Wang K.H., Weitz C., Whitaker C., Wilming L.,  
 RA Wyszynski-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohsaki S.,  
 RA Hayashizaki Y.;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 CC -1- FUNCTION: Stimulator of protein export for NES-containing  
 CC proteins. Also plays a role in the nuclear export of UI snRNA,  
 CC tRNA, and mRNA (By similarity).

CC -1- SUBUNIT: Preferentially binds Ran-GTP. Associates with NXF1, NXF2  
 CC and NXF3.  
 CC -1- SUBCELLULAR LOCATION: Nuclear. Shuttles between the nucleus and  
 CC the cytoplasm.  
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AF156958; AAD54943.2; -  
 DR EMBL: AK003254; BAB2670.1; -  
 DR MGD: MG1:1929619; Nxtl.  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF02136; NTF2; 1.  
 DR PROSITE: PS50177; NTF2\_DOMAIN; 1.  
 KM Transport: Protein transport; mRNA transport; Nuclear protein.  
 FT DOMAIN 16  
 FT CONFID 100 135 NTF2.  
 FT K -> N (IN REF. 2).  
 SQ SEQUENCE 140 AA; 15847 MW; 358EAFDC19FE4594 CRC64;

Alignment Scores:  
 Pred. No.: 1,93e-64 Length: 140  
 Score: 726.00 Matches: 138  
 Percent Similarity: 100.00% Conservative: 2  
 Best Local Similarity: 98.57% Mismatches: 0  
 Query Match: 48.76% Indels: 0  
 DB: 1 Gaps: 0

US-09-763-902b-10 (1-830) x NMT1, MOUSE (1-140)

QY 110 ATGCATCTGTGGATTTCAGACCTATGTGATCAGAGCCGACAGAGCTGCTGAGAGTT 169  
 |||||  
 Db 1 MetabServalasphehystrhyrvalaspelinalcysargalalaaglgluphe 20  
 QY 170 GTCAATGTCTACTACACACCATGATAGCGGCGGCGTGTGTCGCCCTGTACATG 229  
 |||||  
 Db 21 ValasvalatyrtyrthrthmetasplyargrargyleuenseratrgleutyMet 40  
 QY 230 GGCACAGCCACCTGCTGTGATGCGATGCTGTTTCAGACAGCAAGATCTTGAGTGA 289  
 |||||  
 Db 41 GYThralathreulvaltrpaanglyasnalvalaserlygnguslerSerGlu 60  
 QY 290 TTTTGAATGTGGCTTCACAGGAGTCCAAATCAGCGTGTAGATCCAGAGCTGTT 349  
 |||||  
 Db 61 PhephglumelteurProserSerGluPhelGlnIleserValalaspGylnProval 80  
 QY 350 CATGATGAGCCACACACACACACAGCCAGCGTCTGTGTCTACTGTGATCAGTGA 409  
 |||||  
 Db 81 HisaspapalathrProserGlnThrThValleuValallecysglYThVallys 100  
 QY 410 TTTGAGGGGGAACAACAACGGACTTCACAGACTTCATCTGACGCCAGCCCTCA 469  
 |||||  
 Db 101 PheglnglyAsnlysglnarGaspheasnglnasphelleuThrAlaGlnAlaser 120  
 QY 470 CCAGACACACACTGTGGAAGATCCGACAGTACTGCTCCGCTTCACGAGCTGGGCGAC 529  
 |||||  
 Db 121 ProserasnthrValtrplysIlealaseraspGylnaspThrAlaser 140

RESULT 3  
 NMT2\_HUMAN  
 ID NMT2\_HUMAN STANDARD; PRT; 142 AA.  
 AC Q9NP8; Q9NP8; Q9NP8; Q9NP8; Q9NP8; Q9NP8; Q9NP8; Q9NP8; Q9NP8; Q9NP8;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE NTF2-related export protein 2 (p15-2 protein) (DC9) (BM025).  
 GN NMT2.  
 OS Homo sapiens (Human).

OC Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi;  
 OC Mammalia: Eutheria: Primates: Catarrhini: Hominidae: Homo.  
 OX NCBI\_Taxid:9606;  
 RN (1)  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RC TISSUE-Testis;  
 RX MEDLINE:20528640; Pubmed:11073998;  
 RA Herold A., Suyama M., Rodrigues J.P., Braun I.C., Kutay U.,  
 RA Carmona-Fonseca C., Bork P., Izauralde E.,  
 RA "MAP (NXF1) belongs to a multigene family of putative RNA export  
 RT factors with a conserved modular architecture.";  
 RL Mol. Cell. Biol. 23:8996-9008(2000).  
 RN (2)  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RA Kang Y., Cullen B.,  
 RT "p15-2, a homologous protein of p15, interacts with Tap.";  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 RN (3)  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RA Li Y., Peng Y., Li N., Gu W., Han Z., Fu G., Chen Z.,  
 RT "Novel genes expressed in human dendritic cell.";  
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.  
 RN (4)  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE-Ovarian carcinoma;  
 RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,  
 RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,  
 RA Nagatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,  
 RA Takahashi M., Chiba Y., Ishida S., Murakami K., Ono Y., Takiguchi S.,  
 RA Wakamatsu S., Kimura K., Murakami K., Ishii S., Kawai Y., Saito K.,  
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahari K., Masuo Y.,  
 RA Minomiya K., Iwayanagi T.,  
 RT "NEDD human cDNA sequencing project.";  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RN (5)  
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE-Bone marrow;  
 RA Zhao M., Song H., Li N., Peng Y., Han Z., Chen Z.,  
 RT "A novel gene expressed in human bone marrow.";  
 RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.  
 RN (6)  
 RP PARTIAL SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).  
 RA Howden P.,  
 RT Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Regulator of protein export for NES-containing proteins.  
 CC Also plays a role in mRNA nuclear export.  
 CC -1- SUBUNIT: Associates with NXF1, NXF2, NXF3 and NXF5.  
 CC -1- SUBCELLULAR LOCATION: Nuclear. Shuttles between the nucleus and  
 CC the cytoplasm.  
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; A (shown here) and B: are  
 CC produced by alternative splicing.  
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: AJ277591; CAB96371.1; -  
 DR EMBL: AJ278323; CAC01129.1; -  
 DR EMBL: AF246127; AAF78034.1; -  
 DR EMBL: AF201942; AAF86878.1; -  
 DR EMBL: AK023289; BAB14511.1; ALT\_INIT.  
 DR EMBL: AF212223; AAF87325.1; ALT\_INIT.  
 DR EMBL: AL031387; CAB41301.1; -  
 DR EMBL: AL031387; CAB41302.1; -  
 DR MIM: 300320; -  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF02136; NTF2; 1.  
 DR PROSITE: PS50177; NTF2\_DOMAIN; 1.

US-09-763-902B-10 (1-830) x NXL1\_DROME (1-133)

122 GATTTCAGACCTATGTCGATAGCGCCCTCAGACGCTGCGAGGACGTTTGTGATGCTAC 181

111 111 111111 111 111111 111

DB 4 AspleuLYAlaLYsValGluSerCysAlaArgThrAlaSpRhpRhpRhpLeuTYr 23  
 QY 182 TACACCATGATGATAGCGCGCGCTGCTGCTCCCGCTGATGAGGACACCC 241  
 DB 24 TYrAlaSerValaSpaRhpArgGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 43  
 QY 242 CTGGTCTGGAAGGCAATGCTGTTTACAGACAAAGATCTTGAGTAGTTTGAAG 301  
 DB 44 LeuSerTPasngLYasnGlyAlaIleGlyArgGlnMetIleGluSerTYrPheGlnGlu 63  
 QY 302 TTGCTTCAGGAGTTCACAGCGCTGAGTAGAGCTGAGCTGATGATGAGGAGC 361  
 DB 64 LeuProSerSerAsnHisGlnLeuAsnThrLeuAspAlaGlnProIleValaSpGlnAla 83  
 QY 362 ACACCAAGCAGACACCAAGCTGCTGTTGTCATCTGTCAGTGTGAGTGTGAGGAGC 421  
 DB 84 ValSerAsnGlnLeuAlaTYrLeuIleMetAlaSerGlySerValLYsPheAlaSpGln 103  
 QY 422 AAACACGGGACTTCAACCAAGTTCATCTGACCGCCAGCCCTCACCACACACACA 481  
 DB 104 GlnLeuArgLYsPheGlnGlnThrPheIleValThrAlaGlu-----AsnAsp 119  
 QY 482 GTGTGAGAGATGCAAGTACGTCGCTCCGCTCCAGGAC 520  
 DB 120 LysThrLYsValValSerAspCysTYrArgMetGlnGlu 132  
 RESULT 5  
 ID NNT1\_CAEEL STANDARD: PRT: 137 AA.  
 AC 09U757:  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE NTF2-related export protein.  
 GN NNT1 OR Y71F9AM.5.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=20036817; PubMed=10567585;  
 RA Black B.E., Levesque L., Holaska J.M., Wood T.C., Paschal B.M.;  
 RT "Identification of an NTF2-related factor that binds Ran-GTP and  
 RT regulates nuclear protein export.";  
 RL Mol. Cell. Biol. 19:8616-8624(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-Bristol N2;  
 RA Bradshaw-Cordum H., Scott K., Graves T.;  
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: Stimulator of protein export for NES-containing  
 CC proteins. Also plays a role in the nuclear export of. U1 snRNA,  
 CC tRNA, and mRNA (By similarity).  
 CC -1- SUBUNIT: Preferentially binds Ran-GTP (By similarity).  
 CC -1- SUBCELLULAR LOCATION: Nuclear (By similarity).  
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>  
 CC or send an email to [license@isb-sdb.ch](mailto:license@isb-sdb.ch)).  
 CC  
 DR EMBL: AF156960; AAD54945.1;  
 DR EMBL: AC024852; AAK6028.1;  
 DR HSSP: P13662; 10UN.  
 DR WormPep: Y71F9AM.5; CE26780.  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF02136; NTF2.1.  
 DR PROSITE: PS50177; NTF2\_DOMAIN; 1.

KW Transport: Protein transport; mRNA transport; Nuclear protein.  
 FT DOMAIN 19 135 NTF2.  
 SQ SEQUENCE 137 AA; 15479 MW; B99ABE3CA3479962 CRC64;  
 Alignment Scores:  
 Pred. No.: 1,39e-12 Length: 137  
 Score: 206.50 Matches: 40  
 Percent Similarity: 58.59 Conservative: 35  
 Best Local Similarity: 31.25 Mismatches: 48  
 Query Match: 13.878 Indels: 5  
 DB: 1 Gaps: 2  
 US-09-763-902B-10 (1-830) x NNT1\_CAEEL (1-137)  
 QY 140 GATCAGCGCTGACAGACCTCTGAGAGTTGTGTCATGTCTACTACACCATGATAG 199  
 DB 14 GluGlnLeuCYsAsnGlnSerLYsPheMetAspValTYrTYrAspValMetAspArg 33  
 QY 200 CGGCGCGCTTCTGCTCCCGCTGTCATGAGGAGCAGCACCTGCTGGAATGCAAT 259  
 DB 34 LysArgGlnLYsIleIleLYsPheLeuTYrThrGlnValSerAsnAlaValThrPasnGlyAsn 53  
 QY 260 GCTGTTTCAGCAGCAAGAAATCTTGAAGTGAATGTTGCTTCACAGGAGTTC 319  
 DB 54 ProIleAsnGlyTYrAspSerIleCYsGlnPheMetLYsAlaLeuProSerThrGlnHis 73  
 QY 320 CAATCAGCGTGTGACACCTGACCGCTGTCATGATGAAGCCACACCAAGCCAGACC--- 376  
 DB 74 AspIleGlnSerLeuAspAlaGlnArgLeuProGlnGlyValThrGlyAspMetSerGly 93  
 QY 377 ACGTCTCTTGTGTCATCTGTGAGCAGTGTGAGGAGGAGCAACACGGGACTTC 436  
 DB 94 GlyMetLeuLeuAsnValAlaIleGlyAlaValThrValAspLYsPserLYsArgAlaPhe 113  
 QY 437 AACCAAGACTTCATCTGACCGCCAGCGCTCACCAGCAGCAAGTGTGAAGATCGCA 496  
 DB 114 ThrGlnThrLeuLeuLeuLYsValGlu-----AspGlyTYrTYrValLYs 129  
 QY 497 AGTGAAGCTTCCGCTCCAGGAC 520  
 DB 130 SerAspArgPheArgTYrValAsp 137  
 RESULT 6  
 ID NTF2\_HUMAN STANDARD: PRT: 127 AA.  
 AC P13662;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-JAN-1990 (Rel. 13, Last sequence update)  
 DT 15-JUN-2002 (Rel. 41, Last annotation update)  
 DE Nuclear transport factor 2 (NTF-2) (Placental protein 15) (PP15).  
 GN NTF2 OR PP15.  
 OS Homo sapiens (Human).  
 OS Mus musculus (Mouse), and  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OX NCBI\_Taxid=9606, 10090, 10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Human.  
 RA MEDLINE=88247772; PubMed=3380696;  
 RA Grundmann U., Netlich C., Rein T., Lottspeich F., Kuemper H.A.;  
 RT "Isolation of cDNA coding for the placental protein 15 (PP15).";  
 RL Nucleic Acids Res. 16:4721-4721(1988).  
 RN [2]  
 RP SEQUENCE FROM N.A., PARTIAL SEQUENCE, AND CHARACTERIZATION.  
 RC SPECIES-Human.  
 RA MEDLINE=95263674; PubMed=7744965;  
 RA Paschal B.M., Gerace L.;  
 RT "Identification of NTF2, a cytosolic factor for nuclear import that  
 RT interacts with nuclear pore complex protein p62.";  
 RL J. Cell Biol. 129:925-937(1995).  
 RN [3]

PC SPECIES FROM N.A.  
 RA SPECIES-Human, and Mouse;  
 RA Strausberg R.;  
 RL Submitted (FEb-2001) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Mouse: STRAIN-C57BL/6J; TISSUE-Kidney;  
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Aono H., Arai A.,  
 RA Atakawa T., Carlingi P., Fukuda S., Fukunishi Y., Furuno M.,  
 RA Hanagata T., Hara A., Hayatsu N., Hiramoto K., Hirooka T., Hori F.,  
 RA Imetani K., Ishii Y., Itoh M., Izawa M., Kato H., Kawai J., Kojima Y.,  
 RA Kono H., Komda M., Koya S., Kurihara C., Matsuyama T., Miyazaki A.,  
 RA Nishi K., Nomura K., Numazaki R., Ohno M., Okazaki Y., Okido T.,  
 RA Owa C., Saico H., Saico R., Sakai C., Sakai K., Sano H., Sasaki D.,  
 RA Shubaka K., Shubata Y., Shingawa A., Shitaki T., Sogabe Y.,  
 RA Suuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T., Tejima Y.,  
 RA Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
 RA Muramatsu M., Hayashizaki Y.;  
 RA Submitted (Jul-2000) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC SPECIES-Rat; TISSUE-Kidney;  
 RA Kent H.M., Clarkson W.D., Bullock T.L., Stewart M.;  
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS).  
 RC SPECIES-Human;  
 RA MEDLINE-96300274; PubMed-8757804;  
 RX Bullock T.L., Clarkson W.D., Kent H.M., Stewart M.;  
 RT "The 1.6-A resolution crystal structure of nuclear transport factor 2  
 RL (NTF2)".  
 RJ J. Mol. Biol. 260:422-431(1996).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS).  
 RX MEDLINE-98035028; PubMed-9368653;  
 RA Clarkson W.D., Corbett A.H., Paschal B.M., Kent H.M., McCoy A.J.,  
 RA Gerace L., Silver P.A., Stewart M.;  
 RT "Nuclear protein import is decreased by engineered mutants of nuclear  
 RL transport factor 2 (NTF2) that do not bind GDP-Ran".  
 RJ J. Mol. Biol. 272:716-730(1997).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (2.5 ANGSTROMS) OF COMPLEX WITH RAN.  
 RC SPECIES-Rat;  
 RX MEDLINE-98202576; PubMed-9533885;  
 RA Stewart M., Kent H.M., McCoy A.J.;  
 RT "Structural basis for molecular recognition between nuclear transport  
 RL factor 2 (NTF2) and the GDP-bound form of the Ras-family GTPase  
 RJ Ran".  
 RJ J. Mol. Biol. 277:635-646(1998).  
 RN [9]  
 RP FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS.  
 CC INTERACTS WITH THE NUCLEOPORIN P62 AND WITH RAN. ACTS AT A  
 CC RELATIVELY LATE STAGE OF NUCLEAR PROTEIN IMPORT, SUBSEQUENT TO THE  
 CC INITIAL DOCKING OF NUCLEAR IMPORT LIGAND AT THE NUCLEAR ENVELOPE.  
 CC COULD BE PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT  
 CC ASSEMBLE AT THE PORE COMPLEX DURING NUCLEAR IMPORT.  
 CC SUBUNIT: HOMODIMER.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial/  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL, X07315; CAA30278.1; -  
 DR EMBL, X91651; CAA62839.1; -  
 DR EMBL, BC003955; AAH03955.1; -  
 DR EMBL, BC002348; AAH02348.1; -  
 DR EMBL, AK002461; BAB22117.1; -  
 DR EMBL, AK012500; BAB28283.1; -

```

DR EMBL: AK020512: BAB32122.1:
DR EMBL: U43939: AAA85905.1:
DR PIR: S00751: S00751.
DR PDB: 1OUN: 01-APR-97.
DR PDB: 1AR0: 24-DEC-97.
DR PDB: 1ASK: 28-JAN-98.
DR PDB: 1AZK: 29-APR-98.
DR MIM: 605813: -.
DR MGD: MGI:895065; Ntf2.
DR InterPro: IPR002075; NTF2.
DR Pfam: PF02136; NTF2; 1.
DR PROSITE: PS50177; NTF2_DOMAIN; 1.
KW Transport: Protein transport; 3D-structure.
FT DOMAIN 10 NTF2
SQ SEQUENCE 127 AA: 14478 MW: 8177522026262FD3 CRC64:

Alignment Scores:
Pred. No.: 0.000234 Length: 127
Score: 124.00 Matches: 31
Percent Similarity: 43.59% Conservative: 20
Best Local Similarity: 26.50% Mismatches: 54
Query Match: 8.33% Indels: 12
DB: 1 Gaps: 3

US-09-763-902B-10 (1-830) x NTF2_HUMAN (1-127)

QY 167 TTTCGATGTCGTACTACACCCATGATTAAGCGGCGCTTGCTGCCCGCTGTAC 226
Db ||||| ||||| ||||| |||||
QY 14 PheIIcGlnHstYtYrGlnLeuPheAspAsnAspArgThcIcInleuYalaIleTy 33
Db ||||| ||||| ||||| |||||
QY 227 ATGGGCACAGCCACCCCTGCTGCGATGGCATGCTGCTTACAGGACAAGATCCTGAGT 286
Db ||||| ||||| ||||| |||||
QY 34 IleAspAlaSerCysLeuThrTrpGlnGlnGlnInPheGlnInLeuYalaIleVal 53
Db ||||| ||||| ||||| |||||
QY 287 GAGTGTGTTGAATGTCGTCCTTCAGCGAGTCCAAATGAGCTGGTA-----GACTGC 340
Db ||||| ||||| ||||| |||||
QY 54 GlnYsLeuSerSerLeuProheGlnYsIleGlnHISerIleHrAlaGlnAspHis 73
Db ||||| ||||| ||||| |||||
QY 74 GlnProthIhrProAsp-----SerCysIleIleSerMetValIaGly 87
Db ||||| ||||| ||||| |||||
QY 401 TCAAGTGAAGTTGAGGGGACAACAACAGCGGATTAACACAGACTTCATCTGACCGCC 460
Db ||||| ||||| ||||| |||||
QY 88 GlnLeuYAlaAspIhrAspIhrProIleMetGlyPheHisGlnMetPheLeu----- 1050
Db ||||| ||||| ||||| |||||
QY 461 CAGGCGTCACCCAGACACAGATGTCGAGATGCGAAGTGAAGTACGCTCCGC 511
Db ||||| ||||| ||||| |||||
Db 106 -----LysAsnIleAsnAspAlaTrpValCysThrAsnAspMetPheArg 120
Db ||||| ||||| ||||| |||||

RESULT 7
NTF2_XENLA STANDARD; PRT; 127 AA.
AC 042242:
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear transport factor 2 (NTF-2) (P10).
GN NTF2.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipridae; Pipidae;
OC Xenopodinae; Xenopus.
OX NCBI_TaxID=8355;
RN 11)
RP SEQUENCE FROM N.A.
RA Fu R.T., Shetla H.N., Dasso M.;
RL Submitted (SEP-1997) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS.
CC INTERACTS WITH THE NUCLEOPORIN P62 AND WITH RAN. ACTS AT A
CC RELATIVELY LATE STAGE OF NUCLEAR PROTEIN IMPORT. SUBSEQUENT TO THE
CC INITIAL DOCKING OF NUCLEAR IMPORT LIGAND AT THE NUCLEAR ENVELOPE.

```

```

CC CC COULD BE PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT
CC ASSEMBLE AT THE PORE COMPLEX DURING NUCLEAR IMPORT (BY
CC SIMILARITY).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb-sdb.ch).
CC -----
DR EMBL; AF023911; AAB81276.1; -
DR HSSP; P13662; 1A2K.
DR InterPro; IPR002075; NTF2.
DR Pfam; PF02136; NTF2_1.
DR PROSITE; PS50177; NTF2_DOMAIN; 1.
KW Transport; Protein transport.
FT DOMAIN 10 121 NTF2
SO SEQUENCE 127 AA; 14477 MW; 97110D6828268259 CRC64;

Alignment Scores:
Pred. No.: 0.000295 Length: 127
Score: 123.00 Matches: 31
Percent Similarity: 44.44% Conservative: 21
Best Local Similarity: 26.50% Mismatches: 53
Query Match: 8.26% Indels: 12
DB: Gaps: 3

US-09-763-902B-10 (1-830) x NTF2_XENLA (1-127)

QY 167 TTGTGCAATGCTACTACACCAAGTAAAGCGCGCGGTTCGTCCCGCTGTAC 226
Db |||::: ||||| ||| ||| ||| |||::: |||
QY 227 ATGGGCACAGCCACCTTGCTGTGAATGCGAATGCTGTTCAGGACAAAGATCCTTGACT 286
Db |||::: ||| ||| ||| ||| |||::: |||
QY 34 ThraspalaserCysLeuThrTrpDmlGlyngInGlyRnISGLyalaAlaIleVal 53
Db |||::: ||| ||| ||| ||| |||::: |||
QY 287 GAGTGTGTAAGTGGCCCTTCACAGCGAGTTCGCAATCAGCGGTGTA-----GACGTC 340
Db |||::: ||| ||| ||| ||| |||::: |||
QY 54 GlulysLeuSerLeuLeuProPheGlnIlySileGlnHisSerIleThrSerGlnAspHis 73
Db |||::: ||| ||| ||| ||| |||::: |||
QY 341 CAGCGCTGTTCATGATGAAGCCACACACAGCCAGACCGAGTCTGTGTGTCATCTGTGGA 400
Db |||::: ||| ||| ||| ||| |||::: |||
QY 74 GlnProThrProAsp-----SerCysIleIleSerMetValGly 87
Db |||::: ||| ||| ||| ||| |||::: |||
QY 401 TCAATGAGTGTGAGGGAACAAACACGCGACTTCACACCACTTATCTCTACCGCC 460
Db |||::: ||| ||| ||| ||| |||::: |||
QY 88 GlnLeuIlySalaAspAspAspProIleMetGlyPheHisGlnValPheLeuLeu----- 105
Db |||::: ||| ||| ||| ||| |||::: |||
QY 461 CAGCGCTTCACCAACACACAGTGGGAAGATCGCAATGATGATCTTCGCC 511
Db |||::: ||| ||| ||| ||| |||::: |||
QY 106 -----LysaSnIleGlnAspAlaTrpValCysThrAsnAspMetPheArg 120
Db |||::: ||| ||| ||| ||| |||::: |||

RESULT 8
NTF2_CANAL NTF2_CANAL STANDARD: PRT: 124 AA.
AC 09P926:
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Nuclear transport factor 2 (NTF-2).
GN NTF2.
OS Candida albicans (Yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; mitosporic Saccharomycetales; Candida.
OX NCBI_Taxid=5476;
RN [1]
RS SEQUENCE FROM N.A.
RC STRAIN=ATCC 26555;

```

```

Ra Bianchini-Roland S., Cordero-Otero R., Galliardin C., Herrero A.B.,
Ra Dominguez A.,
Rt "Sequence of Candida albicans and Yarrowia lipolytica homolog of the
Rt Saccharomyces cerevisiae NTF2 gene."
Rt Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.
Cc -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS. COULD BE
Cc PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE
Cc AT THE PORE COMPLEX DURING NUCLEAR IMPORT (By SIMILARITY).
Cc -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
Cc -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
Cc -----
Cc Cc This SWISS-PROT entry is copyright. It is produced through a collaboration
Cc between the Swiss Institute of Bioinformatics and the EMBL outstation
Cc the European Bioinformatics Institute. There are no restrictions on its
Cc use by non-profit institutions as long as its content is in no way
Cc modified and this statement is not removed. Usage by and for commercial
Cc entities requires a license agreement (See http://www.isb-sib.ch/announce
Cc or send an email to license@isb-sib.ch).
Cc -----
Cc
Cc EMBL, AF145758; AAF66701.1;
Cc HSSP; P13662; 10UN.
Cc InterPro; IPR002075; NTF2.
Cc Pfam; PF02136; NTF2; 1.
Cc PROSITE; PS50177; NTF2_DOMAIN; 1.
Cc
Cc Kx Transport; Protein transport.
Cc
Cc FT DOMAIN
Cc 8 121 NTF2.
Cc SQ SEQUENCE 124 AA; 14193 MW; B799D1ABA828A049 CRC64;
Cc -----
Cc
Cc Alignment Scores:
Cc
Cc Score: No.: 0 000825 Length: 124
Cc Percent: 118.50 Matches: 40
Cc Percent Similarity: 42.65% Conservative: 18
Cc Best Local Similarity: 29.41% Mismatches: 57
Cc Query Match: 7.96% Indels: 21
Cc DB: 1 Gaps: 6
Cc -----
Cc
Cc US-09-763-902B-10 (1-830) x NTF2_CANAL (1-124)
Cc
Cc QY 116 TCTGTGGATTTCAAGACCTATGTGATCGACGCGCTCGAGCGTGTGAGAGTGTGCAAT 175
Cc Db 2 SerValAspPheHisnla-----ValAlaThrGlnPheCysAsn 14
Cc
Cc QY 176 GTCTACTACACCAACATGATTAAGCGCGCGCTTGTCTGTCCCGCTGTACATGGGCACA 235
Cc Db 15 PheTyrTyrAsnGlnPheAspSerAsparGserGlnLeuGlyAsnLeuTyrArgAsnGlu 34
Cc
Cc QY 236 GCACACCGCTGTGGAAATGGCAATGCTGTTTAGAGCAAGAAGACTCGTAGAGTTT 295
Cc Db 35 SerMetLeuThrPheGlnThrSerGlnLeuGlnGlyAlaArgAspIleValGlnLysLeu 54
Cc
Cc QY 296 GAAATGTGCGCTTCCAGCAG-----TTCCAAATCAGCGTGTGAGTGTCCACCGCTGTT 349
Cc Db 55 AlaSerLeuProPheGlnLysValAlaHisArgIleSerThrIleAspAlaGlnPro--- 73
Cc
Cc QY 350 CATGATGAAGACACACCAAGCCAGACACACCGTCTTGTTCATCTGTGATCA----- 403
Cc Db 74 -----AlaSerAlaAsnGlyAspIleLeuValMetValThrGlyGlnLeuLeu 89
Cc
Cc QY 404 GTGAGATTGGAGGGGACAACAAGAGGAGGAGCTTCACACCAAGAACTTCATCTGACCGCCAG 465
Cc Db 90 IleAspGlnGlnGlnAsnAlaGlnArg--TyrSerGlnValPheHisLeuIle----- 106
Cc
Cc QY 464 GCCTCACCGCAGCAACACAGTGTGGAAGATCGCAAGTGAAGTGAAGTGTCCGC 511
Cc Db 107 -----ProAspAsnGlySerTyrTyrValPheAsnAspIlePheArg 120
Cc
Cc RESULT 9
Cc NTF2_ARATH
Cc ID NTF2_ARATH STANDARD: PRT: 126 AA.
Cc AC Q9C7F5;
Cc
Cc Tr 16-OCT-2001 (Rel. 40, Created)
Cc Tr 16-OCT-2001 (Rel. 40, Last sequence update)
Cc Tr 16-OCT-2001 (Rel. 40, Last annotation update)

```



```

Db      50 Set11val1Ala1ysLeuThrIrrSerLeuProPhagInGInCysLysHSHisIleSerTrn 69
OY      332 GTGATGCGCAGGCTGTATCATATGAAGCACCAAGCACCAAGGACCTTTGTTGTC 391
Db      70 Val1aspGysGInPro-----SerGlyProAlaSerGlyMetLeuValPhe 84
OY      392 ATCTGTGGATCACTGATGATTGAGGGGACCAACAACACGG---GACTTCAACGAGACTTC 448
Db      85 ValSerGlyLysMetLeuGInLeuAlaGlyGInGInH1sAlaLeuLysPheSerGInMetPhe 104
OY      449 ATCTCTG-----ACCGCCCAAGGCTTACACCCAGACACACTGTGGAGATCGCAAGTGAC 502
Db      105 HisLeuMetProThrProGInGInLysSer-----PheTyrVal1PheasnAsp 119
OY      503 TGCTTCGCCG 511
Db      120 IlePheArg 122
RESULT 10
NTF2_YEAST
AC      P33331; STANDARD; PRT; 125 AA.
DT      01-FEB-1994 (Rel. 28, Created)
DT      01-FEB-1995 (Rel. 31, Last sequence update)
DT      16-OCT-2001 (Rel. 40, Last annotation update)
DE      Nuclear transport factor 2 (NTF-2) (Nuclear transport factor p10).
GN      NTF2 OR YER009W.
OS      Saccharomyces cerevisiae (Baker's yeast).
OC      Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC      Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX      NCBI_TaxID=4932.
[1]
RA      SEQUENCE FROM N.A.
RP      STRAIN=528Bc / AB972;
RC      Dietrich F.S., Mulligan J.T., Hennessey K.M., Allen E., Araujo R.,
RA      Aviles E., Berno A., Brennan T., Carpenter J., Chen E., Cherry J.M.,
RA      Chung E., Duncan M., Guzman C., Hartzell G., Hunkeler-Smith S.,
RA      Hyman R., Kayser A., Komp C., LaBkari D., Lew H., Lin D.,
RA      Mosedale D., Nakahara K., Namath A., Norgren R., Oefner P., Oh C.,
RA      Petel F.X., Roberts D., Sehl P., Schramm S., Shogren T., Smith V.,
RA      Taylor P., Wei Y., Yelton M., Bozstein D., Davis R.W.;
RA      Submitted (DEC-1994) to the EMBL/Genbank/DBJ databases.
[2]
RN      1
RP      SEQUENCE OF 1-84 FROM N.A.
RA      Haarer B.K., Petzold A.S., Brown S.S.;
RA      Submitted (JUL-1993) to the EMBL/Genbank/DBJ databases.
[3]
RP      CHARACTERIZATION.
RA      MEDLINE=96178681; Pubmed=860522;
RX      Nehradas U., Blobel G.;
RT      "Role of the nuclear transport factor p10 in nuclear 'import.'";
RL      Science 272:120-122(1996).
CC      -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS.
CC      INTERACTS WITH VARIOUS NUCLEOPORINS AND WITH RAN-GDP. COULD BE
CC      PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE
CC      AT THE PORE COMPLEX DURING NUCLEAR IMPORT. IN VITRO, THE NTF2-RAN-
CC      GDP ASSOCIATION, IN THE PRESENCE OF GTP, TRIGGERS DISSOCIATION OF
CC      THE KARYOPHERIN ALPHA-BETA COMPLEX, ALLOWING NUCLEAR TRANSLOCATION
CC      OF KARYOPHERIN ALPHA AND THE NLS SUBSTRATE.
CC      -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC      -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.
-----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation
CC      at the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC      or send an email to license@isb-sib.ch).
CC      EMBL: U18778; AAB64542.1;
CC      EMBL: L22304; AAB49379.1;
CC      PIR: S41793; S41793.
DR

```

DR HSBP; P13662; LOCN.  
 DR SGP; S0000811; NTF2.  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF01136; NTF2; 1.  
 DR PROSITE: PS50177; NTF2\_DOMAIN; 1.  
 DR TransPort: Protein transport.  
 FT DOMAIN 8 121 NTF2.  
 SO SEQUENCE 125 AA; 14453 MW; 5660D8C00E634714 CRC64;

Alignment Scores:  
 Pred. No.: 0.00131 Length: 125  
 Score: 116.50 Matches: 42  
 Percent Similarity: 44.85% Conservative: 19  
 Best Local Similarity: 30.88% Mismatches: 54  
 Query Match: 7.82% Indels: 21  
 DB: 1 Gaps: 7

US-09-763-902b-10 (1-830) x NTF2\_YEAST (1-125)

OY 116 TCTGTGATTCAGACCTATGTGATCAGCGCTTCAGAGTGTTCAT 175  
 DB 2 SerLeuAspPheAsnThr-----LeuAlaGlnAsnPheThrGln 14  
 OY 176 GTCTACTACACCATGATGATAGCGGCGCTTGTCTGCCGCTGTACATGGCAC 235  
 DB 15 PheTyrTyrAsnGlnPheAspThrAspArgSerGlnLeuGlyAsnLeuTyrArgAsnGln 34  
 OY 236 GCCACCCCTGGTCTGGAATGCAATGCTGTTTCAGACAGAAATCCTTGATGATTTT 295  
 DB 35 SerMetLeuThrPheGlnThrSerGlnLeuGlnGlyAlaTyrAspGlnValGlnLeu 54  
 OY 296 GAAATGTGCTCTCCAGCAGATTCAC-----ATCAGCGTGTGATGCTGCCAGCTGT 349  
 DB 55 ValSerLeuPheProPheGlnValGlnHisArgIleThrThrLeuAspAlaGlnPro--- 73  
 OY 350 CATGTGAGACACCAACCAACGACGACGCTCTGTTGTCATCTGTGATCA----- 403  
 DB 74 -----AlaSerProAsn---GlyAspValLeuValMetIleThrGlyAspLeu 89  
 OY 404 GTGAAGTTTGAAGGGAGCAACACGAGCTTCAACAGACTTATCCTGACGCCGAC 463  
 DB 90 IleAspGlnGlnGlnAsnProGlnArg---PheSerGlnValPheHisIleuLeu----- 106  
 OY 464 GCCTCACCCAGCAACACATGTGGAAGATCGCAATGACTGCTTCCGC 511  
 DB 107 -----ProAspGlyAsnSerTyrTyrValPheAsnAspIlePheArg 120

RESULT 11  
 G3B2\_HUMAN STANDARD: PRT: 482 AA.  
 AC Q9UN86; O75149; O60606; Q90PA1;  
 DT 16-OCT-2001 (Rel. 40; Created)  
 DT 16-OCT-2001 (Rel. 40; Last sequence update)  
 DT 15-JUN-2002 (Rel. 41; Last annotation update)  
 DE Ras GTPase-activating protein binding protein 2 (GAP SH3-domain binding protein 2) (G3BP-2).  
 GN G3BP2 OR KIA0660.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORMS A AND B).  
 RC TISSUE-Brain;  
 RA Kennedy D., Mattick J.S.;  
 RT "Characterisation and chromosomal location of G3BP-1 and G3BP-2a/b, members of a novel SH3 domain-binding and RNA-binding protein family implicated in signal transduction";  
 RT Submitted (Apr-1999) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM A).  
 RC TISSUE-Brain;  
 RX MEDLINE=98403880; PubMed=9734811;

RA Ishikawa K.-I., Nagase T., Suyama M., Miyajima N., Tanaka A.,  
 RA Kotani H., Nomura N., Ohara O.;  
 RT "Prediction of the coding sequences of unidentified human genes. X. The complete sequences of 100 new cDNA clones from brain which can code for large proteins in vitro.";  
 RT DNA Res. 5:169-176(1998).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM B).  
 RC TISSUE-Brain;  
 RA Gultard E.;  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM B).  
 RC TISSUE-Brain;  
 RA Kennedy D., Ru K., Mattick J.S.;  
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORM B).  
 RC TISSUE-B-cell;  
 RA Strusberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP INTERACTION WITH IKAPPABALPHA.  
 RX MEDLINE=20549669; PubMed=10969074;  
 RA Prigent M., Barlat I., Langen H., Dargemont C.;  
 RT "Ikappabalpha and Ikappabalpha/NF-kappa B complexes are retained in the cytoplasm through interaction with a novel partner, RasGAP SH3-binding protein 2.";  
 RL J. Biol. Chem. 275:36441-36449(2000).  
 CC -1- FUNCTION: PROBABLE SCAFFOLD PROTEIN THAT MAY BE INVOLVED IN MRNA TRANSPORT (POTENTIAL).  
 CC -1- SUBUNIT: BINDS TO THE N-TERMINAL DOMAIN OF IKAPPABALPHA.  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; A (SHOWN HERE) AND B; ARE PRODUCED BY ALTERNATIVE SPLICING.  
 CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: AF145284; AAD51932.1; -;  
 DR EMBL: AB014560; BAA31635.1; -;  
 DR EMBL: AF051311; AAC15705.1; -;  
 DR EMBL: AF053535; AAC95292.1; -;  
 DR EMBL: BC011731; AAH11731.1; -;  
 DR HSBP; P09651; IHA1.  
 DR InterPro: IPR002075; NTF2.  
 DR InterPro: IPR005054; RNA\_rec\_mot.  
 DR Pfam: PF00076; itm; 1.  
 DR Pfam: PF02136; NTF2; 1.  
 DR SMART: SM00360; RRM; 1.  
 DR PROSITE: PS50177; NTF2\_DOMAIN; 1.  
 DR PROSITE: PS50102; RRM; 1.  
 DR PROSITE: PS00030; RRM\_RNP\_1; FALSE\_NEG.  
 KW TransPort: RNA-binding; Alternative splicing.  
 FT DOMAIN 11 133 NTF2.  
 FT DOMAIN 134 409 RNA-BINDING (RRM).  
 FT DOMAIN 134 223 GYU-RICH.  
 FT DOMAIN 419 479 GYU-RICH.  
 FT VARSPIC 243 275 MISSING (IN ISOFORM B).  
 FT CONFLICT 267 267 S -> P (IN REF. 2).  
 FT CONFLICT 359 359 E -> V (IN REF. 3).  
 FT CONFLICT 460 460 M -> I (IN REF. 3).  
 SO SEQUENCE 482 AA; 54111 MW; 2545C6A3F1AAE218 CRC64;

Alignment Scores:  
 Pred. No.: 0.00226 Length: 482



Score: 115.50 Matches: 38  
Percent Similarity: 44.09% Conservative: 18  
Best Local Similarity: 29.92% Mismatches: 58  
Query Match: 7.76% Indels: 13  
DB: 1 Gaps: 4

US-09-763-902b-10 (1-830) x G3B2\_HUMAN (1-482)

OY 164 GAGTTTGCAATGCTACTACACACCATGATAGCGGCGGCGTTGCTGTCGCCGCTG 223  
||||| ||||||| :||||| ||| |||  
DB 14 GIupheValaIrgLITyTThLeuLeuAsnLysAlaProGIuTylLeuHHisArgPhe 33  
OY 224 TACATGGGACAGCACCCCTGCTGTAATGCG-----AAT 259  
||||| :||| |||  
DB 34 TyrgIyArgAsnSerTyrValHIsGlyValAlaPalaSerGlyLysProGInGlu 53  
OY 260 GCTGTTTACAGCAAGATCCTTGAGTGAGTTTGAATGCTTCCAGCAGATTC 319  
||||| ||||||| :||| |||  
DB 54 AlavaIlyrgIyGlnAsnAspIleHIsLysValLeuSerLeuAsnPheserGluCys 73  
OY 320 CAATCAGCGTGTGAGTCCAGCGCTGTCATGATGAAGCCACACAGCCAGACGACG 379  
||||| :||| |||  
DB 74 HIsThrLysIle-----ArgHIsValAspAlaHIsAlaThLeuSerAsp---Gly 89  
OY 380 GTCCTGTGTGTCATCTGTGATCAGTGAAGTTTGAGGGCAACAACAGGACTTCAAC 439  
||||| :||| |||  
DB 90 ValValIvalGlnValMetGlyLeuLeuSerAsnSerGlyGlnProGluArgLysPheMet 109  
OY 440 CAGAACTTCATCCCTACCGCCGCGCCAGCCCTACCCACACACAGCTGTGAAGATCCCAAGT 499  
||||| ||||||| :||| |||  
DB 110 GlnThrPheValLeuAlaProGluLysSerValProAsnLysPheTyr---ValHIsAsn 128  
OY 500 GACTGCTCCGCTCCAGGAC 520  
||| |||||||:|||||  
DB 129 AspMetPheArgTyrGluAsp 135

RESULT 12  
G3B2\_MOUSE STANDARD: PRT: 482 AA.

ID G3B2\_MOUSE  
AC P97379: 09R1B8;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Ras-GTPase-activating protein binding protein 2 (GAP SH3-domain binding protein 2) (G3BP-2).  
GN G3BP2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM B).  
RX MEDLINE=98236259; PubMed=9575347;  
RA Kennedy D., Wood S.A., Ramsdale T., Tam P.P., Steiner K.A., Matlick J.S.;  
RT Identification of a mouse orthologue of the human ras-GAP-SH3-domain binding protein and structural confirmation that these proteins contain an RNA recognition motif.";  
RL Biomed. Pept. Proteins Nucleic Acids 2:93-99(1997).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORMS A AND B).  
RA Kennedy D., Matlick J.S.;  
RT Characterisation and chromosomal location of G3BP-1 and G3BP-2a/b, members of a novel SH3 domain-binding and RNA-binding protein family implicated in signal transduction.";  
RT Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: PROBABLE SCAFFOLD PROTEIN THAT MAY BE INVOLVED IN RNA TRANSPORT (POTENTIAL).  
CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS: A (SHOWN HERE) AND B; ARE PRODUCED BY ALTERNATIVE SPLICING.  
CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN.  
CC -1- SIMILARITY: CONTAINS 1 RNA RECOGNITION MOTIF (RRM).  
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation-CC the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial CC entities requires a license agreement (see <http://www.isp-sdb.ch/announce/> or send an email to [license@sdb-sdb.ch](mailto:license@sdb-sdb.ch)).  
CC  
CC EMBL: U65313; AAC53553.1;  
CC EMBL: AF145285; AAD51933.1;  
CC HSSP: P09651; 1HA1; G3BP2.  
CC MGD: MGI:1346339; G3BP2.  
CC InterPro: IPR002075; NTF2.  
CC InterPro: IPR000504; RNA\_rec\_mot.  
CC Pfam: PF00076; rrm; 1.  
CC Pfam: PF02136; NTF2; 1.  
CC SMART: SM00360; RRM; 1.  
CC PROSITE: PS50177; NTF2\_DOMAIN; 1.  
CC PROSITE: PS50102; RRM; 1.  
CC PROSITE: PS00030; RRM\_RNP\_1; FALSE\_NEG.  
CC Transports: RNA-binding; Alternative splicing.  
CC DOMAIN 11 133 NTF2.  
CC DOMAIN 331 409 RNA-BINDING (RRM).  
CC DOMAIN 134 223 GTP-RICH.  
CC DOMAIN 419 479 GLY-RICH.  
CC VARSPLIC 243 275 MISSING (IN ISOFORM B).  
CC CONFLICT 476 482 RETGCR -> TLHRTASLKSCHQMFQWYTLHRVCILWKF  
CC FT FLWMNTOPE (IN REF. 1).  
CC FT F88BC2086123A5CE CRC64;  
CC SO SEQUENCE 482 AA; 54088 MW; F88BC2086123A5CE CRC64;

Alignment Scores:  
Pred. No.: 0.00226 Length: 482  
Score: 115.50 Matches: 38  
Percent Similarity: 44.09% Conservative: 18  
Best Local Similarity: 29.92% Mismatches: 58  
Query Match: 7.76% Indels: 13  
DB: 1 Gaps: 4

US-09-763-902b-10 (1-830) x G3B2\_MOUSE (1-482)

OY 164 GAGTTTGCAATGCTACTACACACCATGATAGCGGCGGCGTTGCTGTCGCCGCTG 223  
||||| ||||||| :||||| ||| |||  
DB 14 GIupheValaIrgLITyTThLeuLeuAsnLysAlaProGIuTylLeuHHisArgPhe 33  
OY 224 TACATGGGACAGCACCCCTGCTGTAATGCG-----AAT 259  
||||| :||| |||  
DB 34 TyrgIyArgAsnSerTyrValHIsGlyValAlaPalaSerGlyLysProGInGlu 53  
OY 260 GCTGTTTACAGCAAGATCCTTGAGTGAGTTTGAATGCTTCCAGCAGATTC 319  
||||| ||||||| :||| |||  
DB 54 AlavaIlyrgIyGlnAsnAspIleHIsLysValLeuSerLeuAsnPheserGluCys 73  
OY 320 CAATCAGCGTGTGAGTCCAGCGCTGTCATGATGAAGCCACACAGGACTTCAAC 379  
||||| :||| |||  
DB 74 HIsThrLysIle-----ArgHIsValAspAlaHIsAlaThLeuSerAsp---Gly 89  
OY 380 GTCCTGTGTGTCATCTGTGATCAGTGAAGTTTGAGGGCAACAACAGGACTTCAAC 439  
||||| :||| |||  
DB 90 ValValIvalGlnValMetGlyLeuLeuSerAsnSerGlyGlnProGluArgLysPheMet 109  
OY 440 CAGAACTTCATCCCTACCGCCGCGCCAGCCCTACCCACACACAGCTGTGAAGATCCCAAGT 499  
||||| ||||||| :||| |||  
DB 110 GlnThrPheValLeuAlaProGluLysSerValProAsnLysPheTyr---ValHIsAsn 128  
OY 500 GACTGCTCCGCTCCAGGAC 520  
||| |||||||:|||||  
DB 129 AspMetPheArgTyrGluAsp 135

RESULT 13  
NTF2\_ORYSA  
ID NTF2\_ORYSA STANDARD: PRT: 122 AA.  
AC Q9XJ54;  
DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Nuclear transport factor 2 (Ntf-2).  
 GN NTF2.  
 OS Oryza sativa (Rice).  
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 CC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 CC Erihartoideae; Oryzaceae; Oryza.  
 ON NCBI\_Taxid=4530;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RC STRAIN=CV. Nipponbare;  
 RA Matsui R., Iwasaki T., Jiang C., Yamamoto N.;  
 RT "Molecular cloning of a cDNA encoding nuclear transport factor 2  
 (Ntf2) from rice."  
 RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS. COULD BE  
 CC PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE  
 CC AT THE PORE COMPLEX DURING NUCLEAR IMPORT (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 DR EMBL: AB011262; BAA81910.1;  
 DR HSSP: P13662; 1A2K  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF02136; NTF2.  
 DR PROSITE: PS50177; NTF2\_DOMAIN. 1.  
 DR KEGG: K01199; NTF2.  
 FT DOMAIN 6  
 FT SEQUENCE 122 AA; 13354 MW; 84100EBB62AB1BD1 CRC64;  
 -----  
 Alignment Scores:  
 Pred. No.: 0.0116 Length: 122  
 Score: 107.00 Matches: 35  
 Percent Similarity: 46.283 Conservative: 21  
 Best Local Similarity: 28.938 Mismatches: 53  
 Query Match: 7.198 Indels: 12  
 Gaps: 4  
 -----  
 US-09-763-902b-10 (1-830) x NTF2\_ORYSA (1-122)  
 OY 158 GCTGAGGAGTTGTCATGCTCTACTACACCAATGATAGCGCGCTTCTCTC 217  
 DB 7 AAlbysAlaPhelAlGluHsYrYrYrPhesPhrAsnaAProAlaLeuVal 26  
 OY 218 CGCCTGATCATGGGACACCGCTGCTGGAATGCTGTTTTCAGGACAGAA 277  
 DB 27 SerLeuYrClnspspGlySerMetLeuThrPheInclnGlnPheLeuGlyAlaAla 46  
 OY 278 TCCCTGAGGAGTTTGTGAATGCTGCTCCAGCAG-----TTCCAATCAGCGTG 331  
 DB 47 AAlaIlAlaGlyLysLeuGlySerLeuProPheAlaInclnYshHsAspIleAsiThr 66  
 OY 332 GTGAGTCCAGCGCTGTTCATGATGAAGACACCAACGACACCGCTCTGTGTC 391  
 DB 67 ValAspCysGlnPro-----SerGlyProGlnGlyMetLeuValPhe 81  
 OY 392 ATCTGGAGTCACTGAGAGTTTGAAGGAGAACAAACA-----CGGAGCTTCAACCGAATTC 448  
 DB 82 ValSerGlySerLeuYrGlnGlyProAspGlnHisProLeuYspSerGlnMetPhe 101  
 OY 449 ATCTCAACGCGCCAGGCTCCACACCAACAGTGTGAGAGTGCAGACTGACTGCTTC 508  
 DB 102 -----GlnLeuLeuProAlaGlyGlyAsnPheryValGlnAsnAspMetPhe 117

OY 509 CGC 511  
 DB 118 Arg 118  
 -----  
 RESULT 14  
 NTF2\_SCHPO STANDARD; PRT; 123 AA.  
 ID NTF2\_SCHPO  
 AC 010100; OSUT4;  
 DT 01-FEB-1996 (Rel. 33, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DE Probable nuclear transport factor 2 (Ntf-2).  
 GN NTF2 OR SPAC1B9.01C OR SPAC15P9.03C.  
 OS Schizosaccharomyces pombe (Fission yeast).  
 CC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC Schizosaccharomycetales; Schizosaccharomycetaceae;  
 CC Schizosaccharomyces.  
 ON NCBI\_Taxid=4896;  
 RN [1]  
 RP SEQUENCE FROM N. A.  
 RC STRAIN=972;  
 RX MEDLINE=21848401; PubMed=11859360;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouros J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,  
 RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagsels K.,  
 RA Jones K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabinowitsch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Voicikert G., Aert R., Robben J., Glynn B.,  
 RA Wellens I., Vanstreels E., Rieger M., Scheer M., Mueller-Auer S.,  
 RA Gadel C., Fuchs M., Fritz C., Holzer E., Moestl D., Hilbert H.,  
 RA Borzjak K., Langer I., Beck A., Leirach H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Medler H., Wambutt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Lelaune V., Motier S.,  
 RA Gallbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garzon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forsburg S.L.,  
 RA Cerrutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovski G.V., Ussery D., Barrett B.G., Nurse P.;  
 RT "The genome sequence of Schizosaccharomyces pombe."  
 RL Nature 415:871-880(2002).  
 CC -1- FUNCTION: FACILITATES PROTEIN TRANSPORT INTO THE NUCLEUS. COULD BE  
 CC PART OF A MULTICOMPONENT SYSTEM OF CYTOSOLIC FACTORS THAT ASSEMBLE  
 CC AT THE PORE COMPLEX DURING NUCLEAR IMPORT (BY SIMILARITY).  
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).  
 CC -1- SIMILARITY: CONTAINS 1 NTF2 DOMAIN  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 -----  
 DR EMBL: AL109951; CAB53052.1;  
 DR EMBL: Z68197; CAA92380.2;  
 DR HSSP: P13662; 1A2K  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF02136; NTF2.  
 DR PROSITE: PS50177; NTF2\_DOMAIN. 1.  
 DR KEGG: K01199; NTF2.  
 FT DOMAIN 7  
 FT SEQUENCE 123 AA; 14078 MW; F587303CA3270E6 CRC64;  
 -----



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: June 24, 2003, 19:25:38 ; Search time 57.1255 Seconds

(without alignments)  
5987.487 Million cell updates/sec

Title: US-09-763-902b-10

Perfect score: 1489

Sequence: 1 tgaacttgatcactgctg.....ataattataatacactg 830

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 1343160

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 0%  
Listing first 45 summaries

Command line parameters:

MODEL=frame\_n2p\_model -DEV=xip  
-O=/cgn2.1/USPO/spool/US09763902/unat\_24062003.130341.12453/bsp\_query.fasta.1.2254  
-DB=SPTRMBL\_21 -GPMT=fastan -SUFFIX=n2p.rspt -MINMATCH=0.1 -LOOPEXT=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blousum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR\_SCORE=pcr -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pro -NOR-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09763902.ecgn.1.1.232-efunat\_24062003.130341.12453 -NCPU=3  
-NO\_MMAP -LARGEROVERY -NEG\_SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database:

SPTRMBL\_21:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phage:\*  
10: sp\_plant:\*  
11: sp\_protent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_virus:\*  
16: sp\_bacteriophage:\*  
17: sp\_archaeal:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	155.5	10.4	167	5	Q8T2G7 dictyostell

2	142	9.5	200	3	Q9HDY6	Q9hdY6 schizosach
3	135.5	9.1	146	10	Q8RUJ2	Q8ruJ2 oryza sativ
4	129	8.7	127	5	Q8R2D5	Q8r2D5 dictyostell
5	125.5	8.4	123	3	Q9P8H0	Q9p8H0 yarrowia 11
6	119	8.0	688	5	Q9I7J5	Q9i7J5 dirosophila
7	119	8.0	690	5	Q9NH72	Q9nh72 dirosophila
8	119	8.0	690	5	Q9VET4	Q9vet4 dirosophila
9	118	7.9	522	10	Q9LPY0	Q9lpY0 dirosophila
10	112.5	7.6	125	3	Q96VN3	Q96vN3 emericella
11	106	7.1	537	5	Q2I3S1	Q2i3S1 caenorhabd
12	105.5	7.1	130	5	Q9VJ85	Q9vj85 dirosophila
13	105	7.1	122	10	Q9FZK4	Q9fzK4 arabidopsis
14	103.5	7.0	796	3	Q9HES7	Q9hes7 kluyveromyc
15	102	6.9	218	4	Q6O599	Q6o599 homo sapien
16	99	6.6	966	10	Q8S0Z3	Q8s0Z3 oryza sativ
17	96	6.4	903	4	Q9UPX1	Q9upX1 homo sapien
18	93	6.2	450	10	Q9FNDO	Q9fnDO arabidopsis
19	92.5	6.2	590	10	Q9LD99	Q9ld99 oryza sativ
20	91	6.1	285	12	Q9YYP6	Q9ypY6 avian adeno
21	90.5	6.1	666	3	Q8TFR8	Q8tfR8 saccharomyc
22	90.5	6.1	666	3	Q8TFR7	Q8tfR7 saccharomyc
23	90.5	6.1	685	3	Q9HGV1	Q9hgV1 saccharomyc
24	90.5	6.1	685	3	Q8TFB8	Q8tfB8 saccharomyc
25	90	6.0	472	11	Q9QX13	Q9qx13 mus musculu
26	88.5	5.9	129	4	Q9HAF6	Q9haF6 homo sapien
27	88.5	5.9	130	5	Q9VRD6	Q9vrD6 dirosophila
28	88.5	5.9	353	17	Q9Y998	Q9y998 aeropyrum p
29	88.5	5.9	507	4	Q96I35	Q96i35 homo sapien
30	88.5	5.9	4823	13	Q93321	Q93321 fuqu rubrip
31	87.5	5.9	460	10	Q9FME2	Q9fmE2 arabidopsis
32	87.5	5.9	629	12	Q9DWH3	Q9dwh3 rat cytomeg
33	87	5.8	485	10	Q9LXL1	Q9lxl1 arabidopsis
34	86.5	5.8	563	13	Q9DDJ4	Q9ddJ4 halichoeres
35	86.5	5.8	719	10	Q8S7D0	Q8s7D0 oryza sativ
36	86.5	5.8	109	10	Q64428	Q64428 chorella v
37	86	5.8	7962	4	Q10465	Q10465 homo sapien
38	86	5.8	34350	4	Q8W422	Q8w422 homo sapien
39	85.5	5.7	147	4	Q9P058	Q9p058 homo sapien
40	85.5	5.7	753	11	Q9JHW8	Q9jhw8 ratuus norv
41	84.5	5.7	872	15	Q9O0Y5	Q9o0Y5 human immun
42	84.5	5.7	926	16	Q92L26	Q92l26 rhizobium m
43	84	5.6	357	6	Q97866	Q97866 sus scrofa
44	84	5.6	362	4	Q9BRD4	Q9brD4 homo sapien
45	84	5.6	499	4	Q8TCP7	Q8tcP7 homo sapien

## ALIGNMENTS

RESULT 1  
Q8T2G7 PRELIMINARY; PRT; 167 AA.  
AC Q8T2G7;  
DT 01-JUN-2002 (TREMBLrel. 21, Created)  
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Hypothetical 19.0 kDa protein.  
OS Dictyostellium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostellida; Dictyostellium;  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HX4;  
RA Gloeckner G., Eichinger L., Szafianski K., Pachbat J., Dear P.,  
RA Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,  
RA Tunggal B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noejei A.A.,  
RT "Sequence and Analysis of Chromosome 2 of Dictyostellium".  
RL Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AC115607; AAL92371.1;  
KW Hypothetical protein.  
SQ SEQUENCE 167 AA; 19039 MW; E16E4A5F4DA0DBAA CRC64;  
Alignment Scores: 6.47e-08 Length: 167  
Pred. No.: 1



```

Db      56 tleuserhneluglnhismetvalalaglialaglugluilegllyatrglyuleuengl 76
QY      301 GTTCCCTTCACGAGCTTCA-----ATCAGCGGTGAGACTGCCAGCTGTTCATGA 354
          |||||  ::  ::  ::  ||||| ||||| |||||
Db      76 yleupropheluglincysatgshialavalcystThyValAarcsylnPro----- 93
QY      355 TGAAGCCACACCAACG---CAAGACACGGTCTTGTTCATCTGGATCGATGAAGTT 411
          ||||| ||||| ||||| ||||| ||||| |||||
Db      94 -----ThrpserpneprogllygllyleuvalPhevalserglyasluendlnh 111
          ||||| ||||| ||||| ||||| ||||| |||||
QY      412 TGAGGGGAGCAAA---CAAGCGAGACTTCAACCAAGAACTCATTCAGCCGCCAGCGCTC 468
          |||  |||  |||  |||  |||  |||  |||
Db      111 ualagllygluglunhsglnleuAlaGpneSerclnmetPheglnleValPrcasngl 131
          |||  |||  |||  |||  |||  |||  |||
QY      469 ACCAGGACACACAGTGTGGAAGATGCAAGTACTGCTTCCGC 511
          |||  |||  |||  |||  |||  |||
Db      131 nglyser-----PhephevalGlinsnspIlePhearg 142
          |||  |||  |||  |||  |||  |||

RESULT 4
ID      Q8RT2D5      PRELIMINARY:      PRT:      127 AA.
AC      Q8RT2D5:
DT      01-JUN-2002 (TREMBLrel. 21, Created)
DT      01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      Putative nuclear transport factor similar to nuclear transport factor
DE      2 (NTP2) from.
OS      Dictyostelium discoideum (Slime mold).
OC      Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.
CX      NCBI_taxid=44689;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN-AX4:
RA      Gloeckner G., Elchinger L., Szafranski K., Pachebat J., Dear P.,
RA      Lehmann R., Baumgart C., Parra G., April J.F., Guigo R., Kumpf K.,
RA      Turgall B., Cox E., Quail M.A., Platzer M., Rosenthal A., Noegel A.A.:
RT      "Sequence and Analysis of Chromosome 2 of Dictyostelium.".
RL      Submitted (APR-2002) to the EMBL/Genbank/DBD databases.
SQ      EMBL, AC115613; AAM10776.1;
DR      SEQUENCE 127 AA; 14317 MW; 2D45B1B20021B298 CRC64;

```

```

Alignment Scores:
Pred. No.: 4.54e-05 127
Score: 129.00 Matches: 33
Percent Similarity: 46.48% Conservative: 33
Best Local Similarity: 23.24% Mismatches: 52
Query Match: 8.66% Indels: 24
DB: 5 Gaps: 5

US-09-763-902B-10 (1-830) x Q8T2D5 (1-127)

QY 95 TCCATAGAACACAGAGATCGCATCTGATTCAGACCTATGTGGATCGAGCCTCGAGA 154
|||:|||||:|||||: |||
Db 3 SerValaapProGlnValValGlyal----- 11

QY 155 GCTGCTGAGAGATTTTGTCAATGTCTACTACACCACTGATAGAGCGGCGCTTGCTG 214
|||||:|||||: ||| ||| |||
Db 12 ---GlyLysGlnPheValGlnIuhIstYrTylGlyIlePheaspSerAsnArgIaGlyLeu 30

QY 215 TCCGCCCTGTACATGAGGGCACACGCCACCTCGTCTGGAATGGCAGTCTGTTACAGACAA 274
|||||:|||||: ||| ||| |||
Db 31 ThrGlnIleYrGlnGlnGlnInThrThrIleuThrItrPgluGlyLysPheLeuSerGlyAla 50

QY 275 GAATCCTTGAGAGGATTTTGAATGTTGGCT-----TCCGCGGAGTTCCAAAACAGC 328
|||||:|||||: ||| ||| |||
Db 51 AspaIaIleValLysHisIleValGlnIubLeuProPheGlnGlnInThrsnArgLysIleAsn 70

QY 329 GTGTAGACTGCCACCTGTCATGATGAAGCCACACAGCCACAGACAGTCTTGTT 388
|||||:|||||: ||| ||| |||
Db 71 SerIleAspCysGlnGlnIthrTyr-----GlnProGlyIleMetIle 84

QY 389 GTCATCTGTGATCACTGATGAGTTTGAGGGGAAACAAACAGGAGTTCACACG---AAC 445
|||:|||||: ||| ||| |||

```

```

Db      85 ThrValThrGlyThrLeuIleIleAspGlyGlyAlaIAs - - - - - AsnGlnLeuIAs 101
QY      446 TTCATCTGACCGCCCGAGCCCTCACCCACACACAGTGTGGATGCCAGTCACTGC 505
      |||:::
      ::: ::::|||
Db      102 PheValGlnAlaPheAsnLeuIleAsaSerAsnAsnGlySerPheLeuLeuIleAsnAspPhe 121
QY      506 TTCGCG 511
      .||||||
Db      122 PheArg 123
      .|||

RESULT 5
Q9P8H0
Q9P8H0 ID Q9P8H0 PRELIMINARY; PRT; 123 AA.
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Rph1.
GN Rph1.
OS Yarrowia lipolytica (Candida lipolytica).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Dipodascaceae; Yarrowia.
OX NCBI_TaxID=4952;
RN [1]
RP SEQUENCE FROM N.A.
RA Cordero Otero R.R., Lepingle A., Galliardin C.;
RT "Rph1 gene from Yarrowia lipolytica."
RL Submitted (MAY-2000) to the EMBL/Genbank/DBJ databases.
  [2]
RP SEQUENCE FROM N.A.
RA Rojas Quijano R., Lepingle A., Gallardin C.;
RL Submitted (APR-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL; AF260231; AAF70316.1; -
DR HSRF; P13662; IOUN.
DR InterPro: IPR002075; NTF2.
DR Pfam: PF02136; NTF2; 1
SQ SEQUENCE 123 AA; 13696 MW; 3E6CB95CFF0058251 CRC64;

```

Alignment Scores:		0.000108	Length:	123
Pred. No.:	Score:	125.50	Matches:	40
Percent Similarity:		44.78%	Conservative:	20
Best Local Similarity:		29.85%	Mismatches:	57
Query Match:		8.43%	Indels:	17
DB:		3	Gaps:	5

US-09-763-902B-10 (1-830) x Q9P8H0 (1-123)

QY	116	TCGTGGATTTCAGACCTATGTGATCGAGCCCTGCAGACSTCTGAGAGATTTCAT	175
Db	2	SevLaApRphasAnThrLeuAlaLysGlnPhcys-----GluPhe-----	15
QY	176	GTCTACTACACACCATGATGAAGCGGCGGCTTCTGCTCCGCGCTGATACAGGCACA	235
Db	16	--TgUTgYgLnThrPheAspThrPaRgSerGlnLeuGlnLeuTgUAdgAspRnIs	34
QY	236	GCCACCCCTGGTGTGGAAATGGCAATGCTGTTTACAGACAGAAACSTTGACATGTTTTT	295
		:::	:::
Db	35	SerMetLeuThrPheThrGlnThrGlnInISgInIAlaInIAlaIleValIGlnLysLeu	54
QY	286	GAAATGTTCTCTCCACGAGC-----TTTSCAAATGACGCGGTAGACCTGCACCGCTT	349
Db	55	ValIGlyLeuPhePheGlyGlnValAlaGlnHisLysIleSerAspIleAspAlaGlnPro--	73
QY	350	CATGATGAAGCCACACCAAGCAGACACGCGCTTGTGTCACTGTGTGATGATAGTGAAG	409
Db	74	-----AlaSerIaGlnIGlyAspValIleValLeuValAlaThrGlnGlyLeuGlyS	90
QY	410	TTTGAGGGGGAACAACACAGGCGACTTCACACAGAACTATCTATGACCGCGGCGCTCA	469
Db	91	ValAspGlyAspAsnProLeuProTgTgGlnValAlaPheHisLeuIle-----	106
QY	470	CCGAGCAACACAGTGTGAAGATCGGAAGTGCATGCTCTCCGC	511

DB 107 PROASglySerSerTyrTyrValPheAsnAspIlePheArg 120

RESULT 6  
ID 091735 PRELIMINARY; PRT: 688 AA.  
AC 091735;  
DT 01-MAR-2001 (TREMBLrel. 16, Created)  
DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Rln. protein.  
GN RIN OR CG9412.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_Taxid=7227;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BERKELEY;  
RX MEDLINE-20196006; PubMed-10731132;  
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
RA Amaralides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
RA Sutton R.G., Mortan J.R., Yandell M.D., Zhang Q., Chen L.X.,  
RA Brandon R.C., Rogers Y.H.C., Blaise R.G., Champe M., Pfeiffer B.D.,  
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Milos G.L.G.,  
RA Abail J.F., Agayani A., An H.-J., Andrews-Planck C., Baldwin D.,  
RA Baller R.M., Basu A., Baxendale J., Bayraktaroglu K., Beasley E.M.,  
RA Beeson K.Y., Bence P.V., Berman B.P., Bhandari D., Bolshakov S.,  
RA Borikova D., Botchan M.R., Bouck J., Broxstein P., Brotler P.,  
RA Burtis K.C., Busan D.A., Butler H., Cadieu E., Center A., Chandra I.,  
RA Cherry J.M., Cawley S., Dahlke C., Dayenport L.B., Davies P.,  
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,  
RA Fostler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
RA Harits N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J.,  
RA Hostin D., Houshon K.A., Howland T.J., Wei M.-H., Ibegam C.,  
RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
RA Mewkow G., Mlshina N.V., Mobarry C., Morris J., Moshrefi A.,  
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
RA Nelson D.R., Nelson K.A., Nixon K.A., Nusskern D.R., Pacleb J.M.,  
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
RA Reluett K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,  
RA Snyder E., Spelling A.C., Stapleton M., Strong R., Sun E.,  
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,  
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,  
RT "The genome sequence of Drosophila melanogaster.";  
RL Science 287:2185-2195(2000).  
DR EMBL: AE003701; AAG22151.1;  
DR FLYBase: FBgn0015778; rln.  
DR InterPro: IPR002075; NTF2.  
DR InterPro: IPR000504; RNA\_rec\_mot.  
DR Pfam: PF002136; NTF2; 1.  
DR Pfam: PF00076; rtm; 1.  
DR SMART: SM00360; RRM; 1.  
DR SMART: PS50102; RRM; 1.  
DR PROSITE: PS50102; RRM; 1.  
SO SEQUENCE 688 AA; 74713 MW; 9E329949E7248C19 CRC64;

Best Local Similarity: 30.33% Mismatches: 54  
Query Match: 7.99% Indels: 12  
DB: 5 Gaps: 5  
US-09-763-902b-10 (1-830) x 091735 (1-688)  
QY 164 GAGTTGTGCAATGTCTACACACACCCATGATGACGGGGGCTTGTGTCGGCTG 223  
DB 17 GlnPheValArgGlnIleTyrThrIleuLeuAsnLysAlaProAsnHisLeuHisArgPhe 36  
QY 224 TACATGGGCACACACACCCCTGTC--TGAAATGCCATGCTGTTTCAGCAAGAATCC 280  
DB 37 TyrAsnHisAsnSerSerTyrIleHisGlyIleuSerLysValValGlyGlnArgLys 56  
QY 281 TTGAGTGAATTTTGAATGTTGCTTCCTCCAGCAG--TTCCAAATCAGCTGGTA 334  
DB 57 IleHisAsnArgIleGlnGlnIleuAsnPheAsnAspCysHisAlaLysIleSerGlnVal 76  
QY 335 GACTGCCACCCCTGTCATGATGACGACACACACACACACGCTCTGTTGTCATC 394  
DB 77 AspAlaGlnAlaThrLeuGlyAsnGly-----ValValValGlnVal 90  
QY 395 TGTGATCAGTGAAGTTGAGGGGACAAACACGAGCTTCAACGAGACTCATCTG 454  
DB 91 ThrGlyGlnLeuSerAsnAspGlyGlnProMetArgArgPheThrGlnIlePheValLeu 110  
QY 455 ACCGCCACGCTCACCACGACACACACGCTGAGATGCAAGTCACTGCTCCGCTTC 514  
DB 111 AlaAlaGln---SerProLysLys-----TyrTyrValHisAsnAspIlePheArgTyr 127  
QY 515 CAGGAC 520  
DB 128 GlnAsp 129  
RESULT 7  
ID 09NH72 PRELIMINARY; PRT: 690 AA.  
AC 09NH72;  
DT 01-OCT-2000 (TREMBLrel. 15, Created)  
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
DE Rasputin.  
GN RIN OR CG9412.  
OS Drosophila melanogaster (Fruit fly).  
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
OC Ephydroidea; Drosophilidae; Drosophila.  
NCBI\_Taxid=7227;  
[1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-20191744; PubMed-10725247;  
RA Pazman C., Maves C.A., Fanto M., Haynes S.R., Mlodzik M.,  
RT "Rasputin, the Drosophila homolog of the RasGAP SH3 binding protein,  
functions in Ras and Rho mediated signalling.";  
RL Development 127:1715-1725(2000).  
DR EMBL: AF231031; AAF68949.1;  
DR FLYBase: FBgn0015778; rln.  
DR InterPro: IPR002075; NTF2.  
DR InterPro: IPR000504; RNA\_rec\_mot.  
DR Pfam: PF02136; rtm; 1.  
DR Pfam: PF00076; rtm; 1.  
DR SMART: SM00360; RRM; 1.  
DR SMART: PS50102; RRM; 1.  
SO SEQUENCE 690 AA; 74985 MW; 40C5BD7B9F5FA253 CRC64;

Alignment Scores:  
Pred. No.: 0.00803 Length: 690  
Score: 119.00 Matches: 37  
Percent Similarity: 45.90% Conservative: 19  
Best Local Similarity: 30.33% Mismatches: 54  
Query Match: 7.99% Indels: 12  
DB: 5 Gaps: 5



US-09-763-902b-10 (1-830) x Q9NH72 (1-690)

QY 164 GAGTTTGTCAATGCTACTACACCAACAGGAGGCGGCGCTTGTCCCGCTG 223  
 |||||  
 19 GIUphValaIrgIntYrTThLeuLeuAsnLysAlaProAsnHISLeuHISArgPhe 38  
 |||||

QY 224 TACATGGGACAGCCACCCGTC---TGGAATGGCAATGCTTTAGACAAGAATCC 280  
 |||||  
 39 TyrAsnHISAsnSerSerTyrIleHISgluSerLysLeuValValGlyGlnArgGlu 58  
 |||||

QY 281 TTGAGTACAGTTTGAAGTGTGCTTCCAGCAG-----TTCCAATCAGCGTGTGA 334  
 |||||  
 59 ILeHISAsnArgIleGlnGlnLeuAsnPhaAsnAspCysHISAlaLysIleSerGlnVal 78  
 |||||

QY 335 GACTGCCAGCGCTTTCATGATGAAGCCACACAGCAGACGAGCTGTGTGTCATC 394  
 |||||  
 79 AspAlaGlnAlaHISLeuGlyAsnGly-----ValValValGlnVal 92  
 |||||

QY 395 TGTGATCACTGAAGTTTGAAGGAGAACAAACAGGACTTCAACAGACTTATCCTG 454  
 |||||  
 93 ThrGlyGluLeuSerAsnAspGlyGlnProMetArgArgPheThrGlnThrPheValLeu 112  
 |||||

QY 455 ACCGCCAGCGCTTCCACCAACACAGTGTGAAGATCGCACTGCTTCCGCTTC 514  
 |||||  
 113 AlaAlaGln---SerProLysLys-----TyrTyrValHISAsnAspIlePheArgTyr 129  
 |||||

QY 515 CAGGAC 520  
 |||||

Db 130 GlnAsp 131

RESULT 8

Q9VFT4 PRELIMINARY: PRT: 690 AA.

AC Q9VFT4  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE CG9412 protein (L031194P).  
 GN RIN OR CG9412.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 NC NCBI\_TaxID=7227;  
 RX MEDLINE=20196006; PubMed=10731132;  
 RC STRAIN-BERKELEY;  
 RP SEQUENCE FROM N.A.  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Goebye J.D.,  
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,  
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,  
 RA Abell J.F., Aghayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Bernier B.P., Bhandari D., Bolshakov S.,  
 RA Butts K.C., Buzam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K.J., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA DuRoi K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Gary N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 RA Jalali M., Kalusz F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,

RA Nelson D.R., Nelson K.A., Nixon K., Nuskern D.R., Pacleb J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,  
 RA Spler E., Spreading A.C., Stapleton M., Strong R., Sun E.,  
 RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zheng G., Zhao Q., Zheng L.,  
 RA Zhang J.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "the genome sequence of Drosophila melanogaster";  
 RL Science 287:2185-2195(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RA Stapleton M., Brokstein P., Hong L., Aghayani A., Carlson J.,  
 RA Champe M., Chavez C., Dorsett V., Farfan D., Frise E., George R.,  
 RA Gonzalez M., Guarin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nuno J., Pacleb J., Paragas V., Park S., Phouanavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celniker S.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AE003701; AAF54965.1;  
 DR EMBL; AY058617; AAL13846.1;  
 DR Flybase; FBgn0015778; rin.  
 DR Interpro; IPR002075; NTF2.  
 DR Interpro; IPR005044; RNA\_rec\_mot.  
 DR Pfam; PF02136; NTF2\_1.  
 DR Pfam; PF00076; rin; 1.  
 DR SMART; SM00360; RRM; 1.  
 DR PROSITE; PS50102; RRM; 1.  
 SQ SEQUENCE 690 AA; 74943 MW; 40C5AC6F9F4EB353 CRC64;

Alignment Scores:

Pred. No.:	0.000803	Length:	690
Score:	119.00	Matches:	37
Percent Similarity:	45.90%	Conservative:	19
Best Local Similarity:	30.33%	Mismatches:	54
Query Match:	7.99%	Indels:	12
DB:	5	Gaps:	5

US-09-763-902b-10 (1-830) x Q9VFT4 (1-690)

QY 164 GAGTTTGTCAATGCTACTACACCAACAGGAGGCGGCGCTTGTCCCGCTG 223  
 |||||  
 19 GIUphValaIrgIntYrTThLeuLeuAsnLysAlaProAsnHISLeuHISArgPhe 38  
 |||||

QY 224 TACATGGGACAGCCACCCGTC---TGGAATGGCAATGCTTTAGACAAGAATCC 280  
 |||||  
 39 TyrAsnHISAsnSerSerTyrIleHISgluSerLysLeuValValGlyGlnArgGlu 58  
 |||||

QY 281 TTGAGTACAGTTTGAAGTGTGCTTCCAGCAG-----TTCCAATCAGCGTGTGA 334  
 |||||  
 59 ILeHISAsnArgIleGlnGlnLeuAsnPhaAsnAspCysHISAlaLysIleSerGlnVal 78  
 |||||

QY 335 GACTGCCAGCGCTTTCATGATGAAGCCACACAGCAGACGAGCTGTGTGTCATC 394  
 |||||  
 79 AspAlaGlnAlaHISLeuGlyAsnGly-----ValValValGlnVal 92  
 |||||

QY 395 TGTGATCACTGAAGTTTGAAGGAGAACAAACAGGACTTCAACAGACTTATCCTG 454  
 |||||  
 93 ThrGlyGluLeuSerAsnAspGlyGlnProMetArgArgPheThrGlnThrPheValLeu 112  
 |||||

QY 455 ACCGCCAGCGCTTCCACCAACACAGTGTGAAGATCGCACTGCTTCCGCTTC 514  
 |||||  
 113 AlaAlaGln---SerProLysLys-----TyrTyrValHISAsnAspIlePheArgTyr 129  
 |||||

QY 515 CAGGAC 520  
 |||||

Db 130 GlnAsp 131

RESULT 9

Q9LPT0



```

ID      Q9LPHY0;   PRELIMINARY;    PRT:       522 AA.
AC      Q9LPHY0;
DT      01-OCT-2000 (TREMBLrel. 15, Created)
DT      01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT      01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE      T73J18.22.
OS      Arabidopsis thaliana (Mouse-ear cress).
OC      Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC      Spermatophytes; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC      eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX      NCBI_TaxId=3702;
RN      [1].
RP      SEQUENCE FROM N.A.
RA      Shinn P., Brooks S., Buehler E., Chao O., Johnson-Hopson C., Khan S.,
RA      Kim G., Alfieri H., Bai Q., Chin C., Chiu J., Choi E., Conn L.,
RA      Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,
RA      Leuz C., Li J., Liu A., Liu K., Liu S., Mukharzky N., Nguyen M.,
RA      Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thavert A.,
RA      Tortumil M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,
RA      Ecker J.R.,
RT      genomic sequence for Arabidopsis thaliana BAC T73J18 from chromosome
RT      1.
RL      Submitted (OCT-2000) to the EMBL/genbank/DDBJ databases.
DR      EMBL; ACCI1661; AAF16635.1; ..
DR      HSSP: P13662; IOUN.
DR      InterPro: IPRO02075; NTF2.
DR      Pfam; PF02136; NTF2; 1.
SQ      SEQUENCE 522 AA; 58484 MW; 74FB3FDFEESBBE6 CRC64;

Alignment Scores:
Pred. No.:           0.000968 . Length:          522
Score:              118.00     Matches:         38
Percent Similarity: 43.44%     Conservative: 15
Best Local Similarity: 31.1%   Mismatches:   49
Query Match:        7.92%      Indels:       20
Db:                10         Gaps:         4

US-09-763-902B-10 (1-830) x Q9LPHY0 (1-522)

OY      98 ATGAACAAGCATGGGATCGTGTGCATTCAAGACCCTATGTGCATCAGGCCCTGCAGACT 157
OY      ||||| |:::|||||
DB      335 lileglucslvalalaserla----- 342
OY      158 GCTGAGAGATTGTGCATGCTCTACACACCACCATGATTAAGCGGCGCGTTGCTGCC 217
OY      ||||| |||||||
DB      343 -----PhelValmslnstlrytlrhlsleupheaspaansprgserSerleuser 359
OY      218 CGCGTGACATGGGCAACCCAGCCCTGGTCGTGAAATGGCAATGCTGTTCAGGCAAGA 277
OY      ||||| ::||::| :||| :||| :||| :||| :||| :|||
DB      360 SerleuylrnsnprothrSerleuleuthrphegluglglnThrllletyglyValasp 379
OY      278 TCCTGTGAGTAGTTTTTAAAATGGTGCCTTCACGCGAGTCCA-----ATCAGCGTG 331
OY      ::::::| ::::| ||||| ::::| |||||
DB      380 AsnlleerasnlyseulysglnleuprophleasprgncysHshlsleullleserThr 395
OY      332 GTAACATGCCAGCCTGTTCATGATGAAGCACACCAACCCAGACCACGCTCTGTGTC 391
OY      ||||| |||||
DB      400 ValaspergrlnPro-----SerSetmetalnaglgylys glylylleuValphe 417
OY      392 ATCTGTGATCATGTGAAGTTGAGGG---AACCAACAAGGCGACTTACCAACAACCTTC 448
OY      ::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
DB      418 ValserclyserlleglnleuhnsiglyuasrhnsproleuarvghesetrglnThrPhe 437
OY      449 ATCTGTG 454
OY      |||
DB      438 Hisleu 439

RESULT 10
Q96VN3 ID#06VNV3 PRELIMINARY; PRT: 125 AA.
NC 01-DEC-2001 (TREMBLrel. 19, Created)
DC 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
```

[illegible]

RP SEQUENCE FROM N.A.  
 RX MEDLINE-99069613; PubMed-9851916;  
 RA none;  
 RT "genome sequence of the nematode *C. elegans*: A platform for  
 investigating biology."  
 RL Science 282:2012-2018(1998).  
 DR EMBL: Z68879; CA93082.1;  
 DR InterPro: IPR002075; NTF2.  
 DR InterPro: IPR005054; RNA-rec.mot.  
 DR Pfam: PF02136; NTF2.1.  
 DR Pfam: PF00076; rim; 1.  
 DR SMART: SM00360; RRM; 1.  
 DR PROSITE: PS0102; RRM; 1.  
 SQ SEQUENCE 537 AA; 59730 MW; 7C3169BFA71C312E CRC64;

Alignment Scores:  
 Pred. No.: 0.0195 Length: 537  
 Score: 106.00 Matches: 30  
 Percent Similarity: 48.03% Conservative: 31  
 Best Local Similarity: 23.62% Mismatches: 48  
 Query Match: 7.12% Indels: 18  
 Gaps: 5

US-09-763-902b-10 (1-830) x Q21351 (1-537)

QY 155 GCTCTGAGAGAG-----TTGTCAATGCTTACTACACACCACCATGATAGCGG 202  
 Db 96 Alalaagluglualglglglalaphacyshtslglnphtyrilthvalserclun 115  
 203 CGGCGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 262  
 Db 116 Argysalalietlrlphtyrilglhslglnserlysphtyriluasaspplnval 135  
 263 GTTTCAGAGACAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 319  
 Db 136 Valnhrdglserlgnlgluilevallylsleutyrnshtslsleuproglnththrlsh 155  
 320 CAATACAGCGTGTAGACTGCCAGCCTGCTCATGATGAGGCACACCAAGCAGCAGC 379  
 Db 156 Lysllelinserrllyslglttyrpro-----ThpRhlslglnl 170  
 380 GTCTTGTGTCATCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 439  
 Db 171 Valletleasnvalilleglthrvlaasleu-----ArpRphleu 185  
 440 CAGACTTCATCTGAGCCGCCAGCCTCACCACAGACAGATGCTGCAATGCTGCAAGT 499  
 Db 186 Glnserpheluleuaglglnglglnglglnglglnglglnglglnglglnglglngl 202  
 500 GACTGCTTCGCGCTTCACAGAC 520  
 Db 203 Aspalahneglnlyrileasp 209

RESULT 12  
 Q9VJ85 PRELIMINARY; PRT; 130 AA.  
 AC Q9VJ85;  
 DT 01-MAY-2000 (TREMblrel. 13, Created)  
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)  
 DT 01-JUN-2001 (TREMblrel. 17, Last annotation update)  
 DE CG10174 protein.  
 GN CG10174.  
 OS Drosophila melanogaster (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;  
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; Drosophila.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BERKELEY;  
 RX MEDLINE-20196006; PubMed-10731132;  
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,  
 Amaratilake P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,

RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,  
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,  
 RA Brandon R.C., Rogers Y.-H.C., Blake J.G., Champagne M., Pfeiffer B.D.,  
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Mkios G.L.G.,  
 RA Abell J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,  
 RA Ballew R.M., Basu A., Baxendale J., Bayraktoglu L., Beasley E.M.,  
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,  
 RA Botkova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,  
 RA Butts K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,  
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,  
 RA de Pablos P., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,  
 RA Dodson K., Dou P.L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,  
 RA Dublin K.J., Evangelista C.C., Ferraz C., Ferlita S., Fleischmann W.,  
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,  
 RA Glodex A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,  
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,  
 RA Hostlin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwan C.,  
 RA Jajuel M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,  
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,  
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,  
 RA Liu X., Matel B., McIntosh T.C., McLeod M.P., McPherson D.,  
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,  
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,  
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclab J.M.,  
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,  
 RA Relbert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,  
 RA Shue B.C., Spidling A.M., Simpson M., Skupski M.P., Smith T.,  
 RA Spletter E., Spradling A.C., Stapleton M., Strong R., Sun E.,  
 RA Svrtskas R., Tector C., Turner R., Ventle E., Wang A.H., Wang X.,  
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,  
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,  
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhen M., Zhang G., Zhao Q., Zheng L.,  
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,  
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;  
 RT "The genome sequence of *Drosophila melanogaster*."  
 RL Science 287:2185-2195(2000).  
 DR EMBL: AE003659; AAF3669.1;  
 DR HSSP: P13662; 100N.  
 DR FlyBase: FBgn0032680; CG10174.  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF02136; NTF2.1.  
 SQ SEQUENCE 130 AA; 14582 MW; 39286FBC5BD7C34E CRC64;

Alignment Scores:  
 Pred. No.: 0.0161 Length: 130  
 Score: 105.50 Matches: 35  
 Percent Similarity: 43.20% Conservative: 19  
 Best Local Similarity: 28.00% Mismatches: 50  
 Query Match: 7.09% Indels: 21  
 Gaps: 6

US-09-763-902b-10 (1-830) x Q9VJ85 (1-130)

QY 161 GAGAGTTTGTCAATGCTTACTACACACCATGAT-----AAGCGCGCGCTTCTG 214  
 Db 12 Lysgluphevalnglglnglglnglglnglglnglglnglglnglglnglglngl 31  
 215 TCCGCCCTGATCAGTGACGACCA-----GCCACCCGCTGCTGATGCAATGCTTCAAG 271  
 Db 32 Ileasnphetrlyrsalalathraspserphetrlrphleglgllyasnlnlglngl 51  
 272 CAGAAATCCTGAGTAGAGTTTGTGAATGTTCCCTCCAGGAGTCCAA----- 322  
 Db 52 Alaprolsyle-----Leuulnysvalnglnserleuserphelglhlysllela 68  
 323 -----ATCAGCGTGTAGACTGCCAGCCTGTTGATGATGAAGCACCACACCAAGCAGACC 376  
 Db 69 Argvallelthrvthvalaspsrlnprrhrseraspglglyl----- 83  
 377 ACGGTCCTGTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 436  
 Db 84 ---ValleullelvalleuaglyargyleuysCysaspaspasprrhrhslalaph 102

QY 437 AACGAGACTTCCTGACCCCGCCGCTCACCACACACAGTGTGGAGATCGCA 496  
 Db 103 SerGlnIlePheLeuLeu-----LysProAsnGlySerLeuPheValala 118  
 QY 497 AGTGACTGCTCCGC 511  
 Db 119 HisAspIlePheArg 123

RESULT 13  
 Q9FZK4  
 ID Q9FZK4 PRELIMINARY: PRT; 122 AA.  
 AC Q9FZK4  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE F17L21.10 (Similar to nuclear transport factor 2).  
 GN F17L21.10.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Johnson-Hopson C., Brooks S., Buehler E., Chao Q., Khan S., Kim C.,  
 RA Shinn P., Altafi H., Bei Q., Chin C., Chlou J., Choi E.,  
 RA Conway A., Gonzales A., Hansen N., Howing B., Koo T., Lam B., Lee J.,  
 RA Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N., Nguyen M.,  
 RA Palm C., Pham P., Sakano H., Schwartz J., Southwick A., Thaveri A.,  
 RA Toriumi M., Vaysberg M., Yu G., Federspiel N.A., Theologis A.,  
 RA Ecker J.R.;  
 RT "Genomic sequence for Arabidopsis thaliana BAC F17L21 from chromosome  
 RT I.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (APR-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (DEC-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RA Ecker J.R.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RA Cheuk R., Shinn P., Brooks S., Buehler E., Chao Q., Johnson-Hopson C.,  
 RA Khan S., Kim C., Altafi H., Bei Q., Chin C., Chlou J., Choi E.,  
 RA Conn L., Conway A., Gonzalez A., Hansen N., Howing B., Koo T., Lam B.,  
 RA Lee J., Lenz C., Li J., Liu A., Liu K., Liu S., Mukharsky N.,  
 RA Nguyen M., Palm C., Pham P., Sakano H., Schwartz J., Southwick A.,  
 RA Thaveri A., Toriumi M., Vaysberg M., Yu G., Davis R., Federspiel N.,  
 RA Theologis A., Ecker J.;  
 RL Submitted (AUG-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
 RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (JUN-2001) to the EMBL/GenBank/DDBJ databases.  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,  
 RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,

RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
 RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
 RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
 RA Ecker J., Theologis A., Davis R.W.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL: AC004557; AAF95749.1;  
 DR EMBL: AY042889; AAK68829.1;  
 DR EMBL: AY072473; AAL68888.1;  
 DR HSP: F13662; IOUN.  
 DR InterPro: IPR002075; NTF2.  
 DR Pfam: PF02136; NTF2; 1.  
 SQ SEQUENCE 122 AA; 13527 MW; E7CDD9486631A1D2 CRC64;

Alignment Scores:  
 Pred. No.: 0.018 Length: 122  
 Score: 105.00 Matches: 33  
 Percent Similarity: 45.45% Conservative: 22  
 Best Local Similarity: 27.27% Mismatches: 54  
 Query Match: 7.05% Indels: 12  
 DB: 10 Gaps: 4

US-09-763-902b-10 (1-830) x Q9FZK4 (1-122)

QY 158 GCTGAGGAGTTGTCAATGTCTACTACACCACCATGGATAAGCGGCGTTGTGCTGC 217  
 Db 7 AlAlYSaLaPheValGluHisTyrSerThrPheAspAlaAsnArgProGlyLeuVal 26  
 QY 218 CGCTGTACATGGCGACACGCCACCTGGTCTCGAATGGCAATGCTCTTCAGGACAAGA 277  
 Db 27 SerLeuTyrGlnGlySerMetLeuThrPheGluGlyGlnLysIleGlnGlySerGln 46  
 QY 278 TCCTTGAGTGAGTGTCTTTTGAATGTGCTCT-----TCCAGCGAGTTCACAAATCAGCGTG 331  
 Db 47 AsnIleValAlaLysLeuThrGlyLeuProPheGlnGlnCysLysHisnIleThr 66  
 QY 332 GTAGACTGCCACCTGTCATGATGAAGCCACCAAGCCAGCCAGCCAGCTCTGTGTGTC 391  
 Db 67 ValAspCysGlnPro-----SerGlyProAlaGlyGlyMetLeuValPhe 81  
 QY 392 ATCTGTGATCAGTGAAGTTTCAGGGGACACAAACACGCG--GACTTCAACCAAGACTTC 448  
 Db 82 ValSerGlyAsnLeuGlnLeuAlaGlyGluGlnHisAlaLeuLysPheSerGlnMetPhe 101  
 QY 449 ATCTGACCGCCAGCCCTCACCAGCAACACAGTGTGGAGATCGCAAGTGAAGTCTTC 508  
 Db 102 HisLeuIleSerAsnGlnGlyAsn-----TyrTyrValPheAsnAspIlePhe 117  
 QY 509 CGC 511  
 Db 118 Arg 118

RESULT 14  
 Q9HFS7  
 ID Q9HFS7 PRELIMINARY: PRT; 796 AA.  
 AC Q9HFS7;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)  
 DE Deubiquitinating enzyme.  
 OS Kluyveromyces lactis (Yeast).  
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;  
 OC Saccharomycetales; Saccharomycetaceae; Kluyveromyces.  
 OX NCBI\_TaxID=28985;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Amerik A., Li S.-J., Hochstrasser M.;  
 RT "Analysis of the deubiquitinating enzymes of the yeast Saccharomyces  
 RT cerevisiae";  
 RL Biol. Chem. 0:0-0(2000).  
 DR EMBL: AF303215; RAG17929.1;  
 DR InterPro: IPR001763; Rhodanese-like.  
 DR InterPro: IPR001394; UCH-2.  
 DR Pfam: PF00442; UCH-1; 1.

DR Pfam: PF00443; UCH-2; 1.  
 DR SMART: SM00450; RHOD; 1.  
 DR PROSITE: PS00972; UCH-2.1; 1.  
 DR PROSITE: PS00973; UCH-2.2; 1.  
 DR PROSITE: PS0235; UCH-2.3; 1.  
 DR SEQUENCE: 796 AA; 90929 MW; 09A9F905579C14B2 CRC64;

## Alignment Scores:

Pred. No.:	Length:	796
Score:	103.50	72
Percent Similarity:	34.53%	Conservative: 34
Best Local Similarity:	23.45%	Mismatches: 122
Query Match:	6.95%	Indels: 79
		Gaps: 14

US-09-763-902b-10 (1-830) x Q9HF57 (1-796)

QY 55 GTTCCCAAGGAGAGAAATACCTGTGGAGCCCTTCATAGAAC-----AGA 108  
 DB 409 lIeValSerSerannglYAsnaAnSerValAlAlYSerGlYhISProSerlIeAnLeu 428  
 QY 109 GATGGCATCTGTGGATTTCAGACCTATGTGATCAGGCGCTGTGAGAGATT 168  
 DB 429 AspPhaAlAlleclYLeuValaAnLeuGlYAsnSerCyStYrLeuAnScys-----lIe 446  
 QY 169 TGTCATGTCTTACACACACCATGATAGCGGCGGCTGTGCTGCCGCTGTACAT 228  
 DB 447 lIleGlncYsLeuLeuGlYCyShISgluLeuSerTrIlehe-LeuThraAnSerTrYar 466  
 QY 229 GGGCAGACAGCCCGCTGTGTAATGGAATGCTGTTCAGACAGAAATCC--TTGAG 285  
 DB 466 glYShISVal-----AsnValaAnSerArYgluGlYSerlYsGlYLeuLeuAl 482  
 QY 286 TGAGTTTTTGAATGTGCTTCAGAGGAGTTCCAAATGACGGTGTAGACTGC--CA 342  
 DB 482 aAnSerYrPheserGlInleuValGlInYsMetYrGlInGlYlYsLeuGlInAlaYrAs 502  
 QY 343 GCCTGTTCATGATGAAGCCACACAGCCAGCCGCTGTTCATCTGTGATC 402  
 DB 502 nAnThrAsMetclYSerThrAlValAlnISProThrGlInPhelYsLeuAlAcYsGlYSe 522  
 QY 403 AGTGAAG-----TTGAGGGGAACAACAAGGAGCTTCAACAG-----AACTTCAT 450  
 DB 522 rIleAsnSerLeuPhelYsGlYglInGlInAlnASpCyGlInlInheCysGlInPhel 542  
 QY 451 CCGTACCGCCAGCGCTCACCACACACAGTGTGGAAGATCCGAAGTACTGCTCCG 510  
 DB 542 ulleuAspGlYLeuHISgluAsnLeuAn-GlncYsGlYlInThraAnProProLeuYsGl 562  
 QY 511 CTGCC-----AGACTGGGCGCACTAGTG 534  
 DB 562 eueSerProgluAlaGlYlYsMetArYglInThraMetProMetArYglleAlaSerAlaIleG 582  
 QY 535 GGGGTGGCAGAGCTCTGTTCATTCAGCCCT 569  
 DB 582 lInrPrGluaYglYrLeuThrThraSpheserValIleValaSpLeuPhelGlInlYnT 602  
 QY 570 --AGCTGTGAGAAATGCAAACTGCACTTCAGAGATGTAGGAACAACAACTTCATT 627  
 DB 602 YrAlaSerGlInleuArYsYlYs----- 609  
 QY 628 TCGTGTTCGGAGACACGACAGCTCC-----ACTG 660  
 DB 610 -----lIeCyAlnAlnIShSerThrYrGlInAlaPheserValleuSerValProV 628  
 QY 661 TGCCGAGGTGAGACTCTTTTGTGCTCAAGTTCTAGAGATCCCTTCTGATATATA 720  
 DB 628 alProArYAlaArYSerCySh-----lIeYrAla 638  
 QY 721 CTGTGTTCATAGTTT-----CCTTTCAAGTAGTAACTTTCTATTTT 768  
 DB 638 spYsPhelYsGlInPhetThrYlYsLeuGlInThrLeuGlInYlYsAspGlInleuTrpYrCySp 658

QY 769 CTACTGCC-----CCATAGAGACTGTATTCGGAATTCGACAAATA 813  
 DB 658 roYrCySlYsGlInArYglInProSerThrYrSglInlellelThrArYgluProAsnA 678  
 QY 814 ATTTAATATACATG 830  
 DB 678 snleullellelHISLeu 683

## RESULT 15

ID	PREDIMINARY:	PRT:	218 AA.
AC 060599;			
DT 01-AUG-1998 (Tremblere, 07, Created)			
DT 01-AUG-1998 (Tremblere, 07, Last sequence update)			
DT 01-AUG-1998 (Tremblere, 07, Last annotation update)			
DE Putative peroxisome microbody protein 175.1.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.			
OX NCBI_TaxId:9606;			
RN [1]			
RP SEQUENCE FROM N.A.			
RA Page N.M., Bullin D.J., Manyonda I., Bicknell A.B., Lowry P.J.;			
RL Submitted (FEB-1998) to the EMBL/GenBank/DBJ databases.			
DR EMBL: AF050199; AAC05749.1;			
SO SEQUENCE 218 AA; 24247 MW; C099F6A676C9BF6F CRC64;			

## Alignment Scores:

Pred. No.:	Length:	218
Score:	102.00	Matches: 54
Percent Similarity:	31.93%	Conservative: 22
Best Local Similarity:	22.69%	Mismatches: 48
Query Match:	6.85%	Indels: 114
		Gaps: 14

US-09-763-902b-10 (1-830) x 060599 (1-218)

QY 24 TCTGTCAACCTTACTTCCTTCAGAGCCCTGTGTCACAGAGAAATAC----- 77  
 DB 40 SerCyShISProTrPrProglYArYrProglYSerProGlInArYhISTrPrLeuHIS 59  
 QY 78 -----CTCGCT-----GGAGCCCTCTTCATAG 101  
 DB 60 CysTrpAspProHISleuGlYglYSerProGlYhISerGlInlYrGlInASpProAla 79  
 QY 102 AACCAAGATGAGCTGTGTGATTCAGACCTATGTGATCAGCGCTGAGACTGCTG 161  
 DB 80 SerTrArYArYsHIS----- 84  
 QY 162 AGGAGTTGTCAATGTCTACTACACACATGATAAGCGGCGGCTTCTGCTCCGCC 221  
 DB 85 -----ProProProTrPrLeuAlaGlYs-----CysProTrPr 95  
 QY 222 TGTCATAGGCGACAG-----CAACCCGCGTGTGGAATGGCAATCTGTTCAG 269  
 DB 96 leuSerTrPrHISglncYsleuLeuIleProPro-----glYGlInThrLeuLeuGlIn 113  
 QY 270 GAC-----AAGATCTTGAAGTATTTTGAAGTGTGCTTCCAGCG 314  
 DB 114 glYhISleuHISAspIleSerAnPro-----LeuPhelLeuArYserGlYSerProVal 131  
 QY 315 AGTTCCAAATCAGCGGTGTGACTGCCAGCTGTTCATGATGAAGCCACACAGCCAGA 374  
 DB 132 TrPrAsnSerCyStr-----LeuMetValIle----- 141  
 QY 375 CCAGGTCCTTGTGTATCTGTGATCAGTGAAGTTTGAAGGGAACAACAACGCGACT 434  
 DB 141 ----- 141  
 QY 435 TCAACAGAACTTCATCTGACGCGCCAGCGCTCACCCAGCAACACAGTGTGAAATCG 494  
 DB 142 PheserArYhISerSerGlInProSerValProValAlaAlaIleleuAla-----Ser 159

Mon Jun 30 08:51:03.2003

us-09-763-902b-10.n2p.rspt

Page 10

```

0Y 455 CAAAGGACGTGGCTTCGGCTTCACGAGACTGGGCGCAGCTAGTGGGGGTGGCAGAGCTCTCTTT 554
      |||:::||||:::|||||
Db 160 ALaalagLuAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAlaAla 172
      :::::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
0Y 555 GCTTCATTACAGCCCTACGCTCTGTAGAGAAATGCAGAACCTCGACTCTCAAGGATGTGAGGA 614
      |||:::|
Db 173 -----CysArg----- 174
0Y 615 ACACAAGTTCATTTCGTGTGTGGGAGACAGCAGCAGCTCCACTGTGGCCAGG 668
      |||||:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
Db 175 -----CysCysCys-----SerSerIleuIuIhArgArgIProArg 186

```

Search completed: June 24, 2003, 19:36:51  
Job time : 64.125 secs

Query Match	85.3%;	Score 978;	DB 10;	Length 1387;
Best Local Similarity	100.0%;	Pred. No. 1.7e-240;		
Matches 978;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0
QY	169	GGATTTTAAACTTATGTGATGATCAGGCATGTGAGACTGCTGAGCAGTTTGCATATATTTA	2228	
Db	75	GGATTTTAAACTTATGTGATGATCAGGCATGTGAGACTGCTGAGGAGTTTGTCAATATTTA	134	
QY	229	CTATGACAAATGATAAAGAAAGACGGGCACTAACCCAGGCTGTATCTGCACAAGGCCAC	2888	
Db	135	CTATGACAAATGATAAAGAAAGACGGGCACTAACCCAGGCTGTATCTGCACAAGGCCAC	194	
QY	289	CTTAATATGGAATGGAATGCTGTTTCAGGGCTGGATGCCCTTAATATATTTTGTGCAC	3488	
Db	195	CTTAATATGGAATGGAATGCTGTTTCAGGGCTGGATGCCCTTAATATATTTTGTGCAC	254	

```

OY 349 ATTGCTTCTAGTGTGATCCAGTCAATATGTATATGCGCAACAGCTTCATGAGCAGC 408
DB 255 ATTGCTTCTAGTGTGATCCAGTCAATATGTATATGCGCAACAGCTTCATGAGCAGC 314
OY 409 AACTCAGTCCCAACTACAGTCTTGTGTGTGACAGCTGGAACCTGGAAGTTGATGAGAA 468
DB 315 AACTCAGTCCCAACTACAGTCTTGTGTGTGACAGCTGGAACCTGGAAGTTGATGAGAA 374
OY 469 CAACACATTTCTTCAACAGCAACTCTCTGTGCTGCTCAGTCCATCTCCACACATATC 528
DB 375 CAACACATTTCTTCAACAGCAACTCTCTGTGCTGCTCAGTCCATCTCCACACATATC 434
OY 529 TGTGTGGAAGATGCAAGTGTGCTGCTGCTTTCAGATGCTGTAGTATGTAAGGGG 588
DB 435 TGTGTGGAAGATGCAAGTGTGCTGCTGCTTTCAGATGCTGTAGTATGTAAGGGG 494
OY 589 CAAGAAGTCCATCTCATTGTGCTCATTAGTTCACAGCAATGGAATTTATGTGAATATTT 648
DB 495 CAAGAAGTCCATCTCATTGTGCTCATTAGTTCACAGCAATGGAATTTATGTGAATATTT 554
OY 649 TGAATGTAGAAAGCACTATATATGTGTGGAACCTAAATTTCTTAAATATTTCTATTCCT 708
DB 555 TGAATGTAGAAAGCACTATATATGTGTGGAACCTAAATTTCTTAAATATTTCTATTCCT 614
OY 709 GTCAGCAGCTTTCTAGCAGCTGCGCAGTTGGAGCATGCGCTCTAAGAGCTTTAAACT 768
DB 615 GTCAGCAGCTTTCTAGCAGCTGCGCAGTTGGAGCATGCGCTCTAAGAGCTTTAAACT 674
OY 769 ATTTTATACAGCTTATATATATATATATATATATATATATATATATATATATATATAT 828
DB 675 ATTTTATACAGCTTATATATATATATATATATATATATATATATATATATATATATAT 734
OY 829 ATGATCTTGTCTAACAATATCTCACTGTGAACCCAGCTATGCAAAATATATATATATAT 888
DB 735 ATGATCTTGTCTAACAATATCTCACTGTGAACCCAGCTATGCAAAATATATATATATAT 794
OY 889 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 948
DB 795 TATATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 854
OY 949 TGGTATATAGGTTAAATTTCTAGTAAACAGCTGCTGCTTTTCAAGTTAACAAGCTGTAAT 1008
DB 855 TGGTATATAGGTTAAATTTCTAGTAAACAGCTGCTGCTTTTCAAGTTAACAAGCTGTAAT 914
OY 1009 GCCATTTTAAATATATAGGCTTTTCAAAATGAGTTCAGTGAATATAGCATTTAGGTTT 1068
DB 915 GCCATTTTAAATATATAGGCTTTTCAAAATGAGTTCAGTGAATATAGCATTTAGGTTT 974
OY 1069 ACATACACTGCTGACATACAGTGAATGATATAGATGTCAGTGGTCTTTTCAAT 1128
DB 975 ACATACACTGCTGACATACAGTGAATGATATAGATGTCAGTGGTCTTTTCAAT 1034
OY 1129 TTAAGTAAATTTTGTGTTT 1146
DB 1035 TTAAGTAAATTTTGTGTTT 1052

RESULT 2
US-09-764-864-1682/c
Sequence 1682, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT223
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1682
LENGTH: 8918
TYPE: DNA

```

```

ORGANISM: Homo sapiens
US-09-764-864-1682
Query Match 65.2% Score 747.2; DB 10; Length 8918;
Best Local Similarity 99.6%; Pred. No. 7e-181;
Matches 749; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

OY 395 GTTCATGAGCAAGCACTCAGTCCCAACTCAAGTCTTGTGTGACAGTCAAGCTGTG 454
DB 2199 GATGAGAGCAAGCACTCAGTCCCAACTCAAGTCTTGTGTGACAGTCAAGCTGTG 2140
OY 455 AAGTTGATGGAACAACAACATTTCTTCAACAGCAACTCTGCTGCTGCTGCTGCTGCTGCT 514
DB 2139 AAGTTGATGGAACAACAACATTTCTTCAACAGCAACTCTGCTGCTGCTGCTGCTGCTGCT 2080
OY 515 ACTCCCAACAATCTGTGTGGAAGATGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 574
DB 2079 ACTCCCAACAATCTGTGTGGAAGATGCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2020
OY 575 AGTAGTAAAGGGGCAAAAGTCCATCTTGTGCTCATTAGTTCAGCAATGGAAT 634
DB 2019 AGTAGTAAAGGGGCAAAAGTCCATCTTGTGCTCATTAGTTCAGCAATGGAAT 1960
OY 635 TATGGAATTTATTTGATTGTAGAGCACTATATATGCTGGAACCTAAATTTCTTTAA 694
DB 1959 TATGGAATTTATTTGATTGTAGAGCACTATATATGCTGGAACCTAAATTTCTTTAA 1900
OY 695 TATTTTCAATCTGCTGACAGCACTTTCTAGCAGCTGCGCAGTTGGAGCATTGCCCTCTTA 754
DB 1899 TATTTTCAATCTGCTGACAGCACTTTCTAGCAGCTGCGCAGTTGGAGCATTGCCCTCTTA 1840
OY 755 AGAGCTTAAACCTATTTTATACATGCTTATATACATTCACATATGACATTTCTTTAA 814
DB 1839 AGAGCTTAAACCTATTTTATACATGCTTATATACATTCACATATGACATTTCTTTAA 1780
OY 815 ATATATATTAACACATGATCTGTGTAACATATCTCACTGTAACCCAGCTATGGA 874
DB 1779 ATATATATTAACACATGATCTGTGTAACATATCTCACTGTAACCCAGCTATGGA 1720
OY 875 AATATATCTTTTATATATATATATATATATATATATATATATATATATATATATATAT 934
DB 1719 AATATATCTTTTATATATATATATATATATATATATATATATATATATATATATAT 1660
OY 935 GAAGTGAAGTTTGTGTTATAGGTTAATTTCTAGTAAACAGCTGCTGTTTTCG 994
DB 1659 GAAGTGAAGTTTGTGTTATAGGTTAATTTCTAGTAAACAGCTGCTGTTTTCG 1600
OY 995 TTAACACTGTGTAATCCATTTTATATATATAGCTTTTCAATCAGTTCAGTGAATAATG 1054
DB 1599 TTAACACTGTGTAATCCATTTTATATATATAGCTTTTCAATCAGTTCAGTGAATAATG 1540
OY 1055 ACAGATTTAGGTTTACATATCTGACATACAGTGAATGCAATATAGAGATGTTCACT 1114
DB 1539 ACAGATTTAGGTTTACATATCTGACATACAGTGAATGCAATATAGAGATGTTCACT 1480
OY 1115 GGTGCTTTTCAATTTTAAATATTTTGTGTTT 1146
DB 1479 GGTGCTTTTCAATTTTAAATATTTTGTGTTT 1448

RESULT 3
US-09-764-864-1683/c
Sequence 1683, Application US/09764864
Patent No. US20020132753A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies
FILE REFERENCE: PT223
CURRENT APPLICATION NUMBER: US/09/764,864
CURRENT FILING DATE: 2001-01-17
Prior application data removed - consult PALM or file wrapper
NUMBER OF SEQ ID NOS: 1792
SOFTWARE: PatentIn Ver. 2.0

```

```

RESULT 4
US-09-796-692-8289
; Sequence 8289, Application US/09796692
; Publication No. US20020198362a1
; GENERAL INFORMATION:
; APPLICANT: Gaiger, Alexander
; APPLICANT: Algate, Paul A.
; APPLICANT: Mannion, Jane
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THERAPY OF
; TITLE OF INVENTION: HEMATOLOGICAL MALIGNANCIES

```

RESULT 5  
US-10-040-862-8289  
; Sequence 8289, Application US/10040862  
; Publication No. US20030078396A1  
; GENERAL INFORMATION:



APPLICANT: Galger, Alexander  
 APPLICANT: Algate, Paul A.  
 APPLICANT: Mannion, Jane  
 APPLICANT: Reltter, Marc  
 APPLICANT: Corixa Corporation  
 TITLE OF INVENTION: Compositions and Methods for the Detection, Diagnosis and Therapy  
 TITLE OF INVENTION: Hemtological Malignancies  
 FILE REFERENCE: 014058-01352005  
 CURRENT APPLICATION NUMBER: US/10/040,862  
 CURRENT FILING DATE: 2001-11-06  
 PRIOR APPLICATION NUMBER: US 60/186,126  
 PRIOR FILING DATE: 2000-03-01  
 PRIOR APPLICATION NUMBER: US 60/190,479  
 PRIOR FILING DATE: 2000-03-17  
 PRIOR APPLICATION NUMBER: US 60/200,545  
 PRIOR FILING DATE: 2000-04-27  
 PRIOR APPLICATION NUMBER: US 60/200,303  
 PRIOR FILING DATE: 2000-04-28  
 PRIOR APPLICATION NUMBER: US 60/200,779  
 PRIOR FILING DATE: 2000-04-28  
 PRIOR APPLICATION NUMBER: US 60/200,999  
 PRIOR FILING DATE: 2000-05-01  
 PRIOR APPLICATION NUMBER: US 60/202,084  
 PRIOR FILING DATE: 2000-05-04  
 PRIOR APPLICATION NUMBER: US 60/206,201  
 PRIOR FILING DATE: 2000-05-22  
 PRIOR APPLICATION NUMBER: US 60/218,950  
 PRIOR FILING DATE: 2000-07-14  
 PRIOR APPLICATION NUMBER: US 60/222,903  
 PRIOR FILING DATE: 2000-08-03  
 PRIOR APPLICATION NUMBER: US 60/223,416  
 PRIOR FILING DATE: 2000-08-04  
 PRIOR APPLICATION NUMBER: US 60/223,378  
 PRIOR FILING DATE: 2000-08-07  
 PRIOR APPLICATION NUMBER: US 09/796,692  
 PRIOR FILING DATE: 2001-03-01  
 NUMBER OF SEQ ID NOS: 10467  
 SOFTWARE: FastSeq for Windows Version 3.0  
 SEQ ID NO 8289  
 LENGTH: 464  
 TYPE: DNA  
 ORGANISM: Homo sapiens  
 US-10-040-862-8289

Query Match: 36.2%; Score 414.8; DB 9; Length 464;  
 Best Local Similarity: 99.5%; Pred. No. 2.1e-96;  
 Matches 416; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

729 CTGCCAGTTGGAGCATTGCCCTCTAAGAGCTTTAAACATATTTTACATGCTTATA 788  
 1 CTGCCAGTTGGAGCATTGCCCTCTAAGAGCTTTAAACATATTTTACATGCTTATA 60  
 789 TACATTCACATTAAGACATTTCTTAATAATTAACATGATCTGTACTAACAATA 848  
 61 TACATTCACATTAAGACATTTCTTAATAATTAACATGATCTGTACTAACAATA 120  
 849 CTCACGTGAAACCCAGCCTATTGCAAAATATTAATATTAATATCTATGGAT 908  
 121 CTCACGTGAAACCCAGCCTATTGCAAAATATTAATATTAATATCTATGGAT 180  
 909 CTCACGTGAAACCCAGCCTATTGCAAAATATTAATATTAATATCTATGGAT 968  
 181 CTCACGTGAAACCCAGCCTATTGCAAAATATTAATATTAATATCTATGGAT 240  
 969 TAGTAAACAGTTCCTGTTTCTGTTAAGACCTGTATGCAATTAATATGAGCT 1028  
 241 TAGTAAACAGTTCCTGTTTCTGTTAAGACCTGTATGCAATTAATATGAGCT 300  
 1029 TTTCAAATCAGTTCAGTAAATATAGTACAGATTTAGTTTACATAGTCTGACATAC 1088  
 301 TTTCAAATCAGTTCAGTAAATATAGTACAGATTTAGTTTACATAGTCTGACATAC 360  
 1089 TGGAAATGCATATAGAGATGTTCAAGTGTCTTTTCAATTTTAAAGTAATTTTGT 1146

DB 361 TGGAAATGCATATAGAGATGTTCAAGTGTCTTTTCAATTTTAAAGTAATTTTGT 418

RESULT 6  
 US-09-764-864-394  
 Sequence 394, Application US/09764864  
 Patent No. US20020132753A1  
 GENERAL INFORMATION:  
 APPLICANT: Rosen et al.  
 TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 FILE REFERENCE: PT23  
 CURRENT APPLICATION NUMBER: US/09/764,864  
 PRIOR APPLICATION data removed - consult PAM or file wrapper  
 NUMBER OF SEQ ID NOS: 1792  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 394  
 LENGTH: 864  
 TYPE: DNA

ORGANISM: Homo sapiens  
 FEATURE:  
 NAME/KEY: SITE  
 LOCATION: (1)  
 OTHER INFORMATION: n equals a,t,g, or c  
 NAME/KEY: SITE  
 LOCATION: (7)  
 OTHER INFORMATION: n equals a,t,g, or c  
 NAME/KEY: SITE  
 LOCATION: (15)  
 OTHER INFORMATION: n equals a,t,g, or c  
 NAME/KEY: SITE  
 LOCATION: (9)  
 OTHER INFORMATION: n equals a,t,g, or c  
 NAME/KEY: SITE  
 LOCATION: (837)  
 OTHER INFORMATION: n equals a,t,g, or c  
 NAME/KEY: SITE  
 LOCATION: (853)  
 OTHER INFORMATION: n equals a,t,g, or c  
 NAME/KEY: SITE  
 LOCATION: (863)  
 OTHER INFORMATION: n equals a,t,g, or c  
 US-09-764-864-394

Query Match: 22.0%; Score 251.6; DB 10; Length 864;  
 Best Local Similarity: 72.4%; Pred. No. 2.1e-54;  
 Matches 326; Conservative 0; Mismatches 124; Indels 0; Gaps 0;

152 GCTCTCTGATTTTAAACCTTATGATGATGATGATGATGATGATGATGATGATGAT 221  
 79 CATCTGTGATTTTCAAGACCTATGATGATGATGATGATGATGATGATGATGATGAT 138  
 222 ATATTACTAGACCAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 281  
 139 ATGTCTACTACACCAACATGATGATGATGATGATGATGATGATGATGATGATGAT 198  
 282 AGCCACCTTATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 341  
 199 CAGCCACCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 258  
 342 TTGACACATGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 401  
 259 TTGAATGTGCTTTCAGCCAGTTCATATGATGATGATGATGATGATGATGATGATGAT 318  
 402 AGCAACACTAGTCCCAAACTACAGTTCCTGTTGATGATGATGATGATGATGATGAT 461  
 319 ATGAAGCCACCAACCAAGCCAGGATGATGATGATGATGATGATGATGATGATGAT 378  
 462 ATGAAGCCACCAACCAATTTCTTCAACCAAGACTCTCTGATGATGATGATGAT 521

Db 379 AGGGAAACAACAGGAGCTCAACAGACTCATCTGACCGCCAGCCAGCCACCA 438  
|||  
Qy 522 ACAATCTGTGTGGAATATGCAAGTATGCTCCCTTTTCAAGATGGCTGTAGT 581  
|||  
Db 439 GCAACACAGTGTGGAATATGCAAGTATGCTCCCTTTTCAAGACTGGCGTAGT 498  
|||  
Qy 582 AAAGGGCAAAAGTCTCTTCATTTGTC 611  
|||  
Db 499 GGGGGTGGCAGAGTCTCTTTCATTC 528  
|||

## RESULT 7

US-09-815-343-1201  
Sequence 1201, Application US/09815343  
Patent No. US2001005596A1  
GENERAL INFORMATION:  
APPLICANT: Meagher, Madeleine  
APPLICANT: Xu, Jiangchun  
APPLICANT: King, Gordon E.  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND  
FILE REFERENCE: 210121.504  
CURRENT APPLICATION NUMBER: US/09/815, 343  
CURRENT FILING DATE: 2001-03-22  
NUMBER OF SEQ ID NOS: 1556  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 1201  
LENGTH: 318  
TYPE: DNA  
ORGANISM: Homo sapien  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: (1)...(318)  
OTHER INFORMATION: n = A,T,C or G  
US-09-815-343-1201

Query Match 16.1% Score 184.8; DB 10; Length 318;  
Best Local Similarity 73.8%; Pred. No. 1.6e-37;  
Matches 234; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Qy 195 CATGTAGAGCTGCTGAGAGTTTGTCAATATTACTATGACATGATGATAAAGAC 254  
|||  
Db 1 CCTGCAAGCTGCTGAGAGTTTGTCAATATTACTATGACATGATGATAAAGCGGC 60  
|||  
Qy 255 GGGCACTAACAGGCTGTATCTGACAAAGCCACCTTAATATGGAATGGAATGCTTT 314  
|||  
Db 61 GTTGTCTGCTCCGCTGTACATGGGACAGCCACCTGTGTGGAATGGAATGCTTT 120  
|||  
Qy 315 CAGGCTGAGATGCCCTTAATATTATTTTGAACATTCCTTGAATGAGTCCAGTCA 374  
|||  
Db 121 CAGGACAAAGATCTTGAATGAGTTCCTTGAATGAGTTCCTTGAATGAGTTCCTTGA 180  
|||  
Qy 375 ATATGTAGATTTGCAACAGCTTCAATGAGCAAGCACTACCTCAACTACAGTCTTG 434  
|||  
Db 181 GCGTGTAGTACTGCGAGCTTTCATGATGAAGCACCACCAAGCCAGACAGGCTCTTG 240  
|||  
Qy 435 TTGTGACCAAGTGAAGTGTGAATTTGATGGAACAAACAACTTTTCAACAGAACT 494  
|||  
Db 241 TTGTGATCTGTGATGATGATGAATTTGAGGGGAACAAACAAAGGAGCTTCAACGAGACT 300  
|||  
Qy 495 TCCTGCTGACTGCTCAG 511  
|||  
Db 301 TCATCTGACCGCCAG 317  
|||

## RESULT 8

US-09-796-692-4160  
Sequence 4160, Application US/09796692  
Publication No. US20020198362A1  
GENERAL INFORMATION:  
APPLICANT: Galger, Alexander  
APPLICANT: Algate, Paul A.

APPLICANT: Mannion, Jane  
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE DETECTION, DIAGNOSIS AND THER  
FILE REFERENCE: 2077 001200  
CURRENT APPLICATION NUMBER: US/09/796, 692  
CURRENT FILING DATE: 2001-03-01  
PRIOR APPLICATION NUMBER: 60/186,126  
PRIOR FILING DATE: 2000-03-01  
PRIOR APPLICATION NUMBER: 60/150,479  
PRIOR FILING DATE: 2000-03-17  
PRIOR APPLICATION NUMBER: 60/200,545  
PRIOR FILING DATE: 2000-04-27  
PRIOR APPLICATION NUMBER: 60/200,303  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,779  
PRIOR FILING DATE: 2000-04-28  
PRIOR APPLICATION NUMBER: 60/200,999  
PRIOR FILING DATE: 2000-05-01  
PRIOR APPLICATION NUMBER: 60/202,084  
PRIOR FILING DATE: 2000-05-04  
PRIOR APPLICATION NUMBER: 60/206,201  
PRIOR FILING DATE: 2000-05-22  
PRIOR APPLICATION NUMBER: 60/218,950  
PRIOR FILING DATE: 2000-07-14  
PRIOR APPLICATION NUMBER: 60/222,903  
PRIOR FILING DATE: 2000-08-03  
PRIOR APPLICATION NUMBER: 60/223,416  
PRIOR FILING DATE: 2000-08-04  
PRIOR APPLICATION NUMBER: 60/223,378  
PRIOR FILING DATE: 2000-08-07  
NUMBER OF SEQ ID NOS: 9597  
SOFTWARE: FastSeq for Windows Version 3.0  
SEQ ID NO 4160  
LENGTH: 318  
TYPE: DNA  
ORGANISM: Homo sapiens  
US-09-796-692-4160

Query Match 16.1% Score 184.2; DB 9; Length 318;  
Best Local Similarity 73.8%; Pred. No. 2.2e-37;  
Matches 234; Conservative 0; Mismatches 83; Indels 0; Gaps 0;

Qy 195 CATGTAGAGCTGCTGAGAGTTTGTCAATATTACTATGACATGATGATAAAGAC 254  
|||  
Db 1 CCTGCAAGCTGCTGAGAGTTTGTCAATATTACTATGACATGATGATAAAGCGGC 60  
|||  
Qy 255 GGGCACTAACAGGCTGTATCTGACAAAGCCACCTTAATATGGAATGGAATGCTTT 314  
|||  
Db 61 GTTGTCTGCTCCGCTGTACATGGGACAGCCACCTGTGTGGAATGGAATGCTTT 120  
|||  
Qy 315 CAGGCTGAGATGCCCTTAATATTATTTTGAACATTCCTTGAATGAGTCCAGTCA 374  
|||  
Db 121 CAGGACAAAGATCTTGAATGAGTTCCTTGAATGAGTTCCTTGAATGAGTTCCTTGA 180  
|||  
Qy 375 ATATGTAGATTTGCAACAGCTTCAATGAGCAAGCACTACCTCAACTACAGTCTTG 434  
|||  
Db 181 GCGTGTAGTACTGCGAGCTTTCATGATGAAGCACCACCAAGCCAGACAGGCTCTTG 240  
|||  
Qy 435 TTGTGACCAAGTGAAGTGTGAATTTGATGGAACAAACAACTTTTCAACAGAACT 494  
|||  
Db 241 TTGTGATCTGTGATGATGATGAATTTGAGGGGAACAAACAAAGGAGCTTCAACGAGACT 300  
|||  
Qy 495 TCCTGCTGACTGCTCAG 511  
|||  
Db 301 TCATCTGACCGCCAG 317  
|||

## RESULT 9

US-10-040-862-4160  
Sequence 4160, Application US/10040862  
Publication No. US20030078396A1  
GENERAL INFORMATION:  
APPLICANT: Galger, Alexander



[illegible]

APPLICANT: BERLIN, Kurt  
 TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation  
 FILE REFERENCE: 5013.1003  
 CURRENT APPLICATION NUMBER: US/10/239,676  
 CURRENT FILING DATE: 2002-09-24  
 PRIOR APPLICATION NUMBER: PCT/EP01/03968  
 DE 10019058.8  
 DE 10019173.8  
 DE 10032529.7  
 DE 10043826.1  
 PRIOR FILING DATE: 2001-04-06  
 2000-04-06  
 2000-04-07  
 2000-06-30  
 2000-09-01  
 NUMBER OF SEQ ID NOS: 228  
 SEQ ID NO 90  
 LENGTH: 5689  
 TYPE: DNA  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)  
 FEATURE:  
 NAME/KEY: unsure  
 LOCATION: 1818, 1930..1931, 1934..1935, 1947..1948, 1966)  
 US-10-239-676-90

	Query Match	Similarity	3.8%;	Score 43;	DB 9;	Length 5669;
	Best Local	Similarity	44.5%;	Pred. No. 1.8;		
	Matches	200;	Conservative	0;	Mismatches 243;	Indels 6; Gaps 1;
QY	524	AATACCTGTGGAAGATTCAGTCAATTCCTCCGTTTCAAGATTTGGCTAGATTAA	583			
Db	1610	ATTATATTTTCATGATATTAATGTTTTTGGTTGTATTATTAATGAATTAAT	1669			
QY	584	AGGGGCAAAAGTCCTCTCATTTGGTCATTAGTCCAGCAATTTGAATTTATGTAAT	643			
Db	1670	ATTGCGATATTTTATTTGTATTAATTTTTCGTTTTTTTATTTTAAATTTTTTATTT	1729			
QY	644	TATTTGATTTGTAGACACATTAATATATGCTGAACATAATTTCTTATATTTTCTA	703			
Db	1730	TTTTTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTT	1788			
QY	704	TTCTGTGACAGCCCTTTCTAGACAGTCCAGTTGGAGCAATGCCCCCTAAGACTTA	763			
Db	1790	TTTTTATTTATTTTATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTATTT	1843			
QY	764	AAACTATTTTATTTACATGCTTAATATACATTTCCACTAATGACATTTTATATATTA	823			
Db	1844	AAAAAATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT	1903			
QY	824	AACGACATGATCTGGTACTACATACACTGACAGTGAACCCAGTCATGCCAATAATTA	883			
Db	1904	TTTATATTTATTTATTTATTTTATTTTATTTTATTTTATTTTATTTTATTTTATTTT	1963			
QY	884	CTTTTATATATATTTATCTATGATGTCAGACATATACACTCTGGGAAGATGAG	943			
Db	1964	ATNTTTTTTTTATTTATTTTATTTTAAAAATTTTTTTTTTTTTTTTTTTTTTTTTTT	2023			
QY	944	TTTTTTGGTATTTAGTTAATTTTCTACT	972			
Db	2024	TTTTTTTTTTTTTATTTGTTTTTTTTTATTTATTT	2052			

RESULT 14  
US-10-239-676-209  
; Sequence 209, Application US/10239676  
; Publication No. US20030082609a1  
; GENERAL INFORMATION:  
; APPLICANT: OLEK, Alexander  
; APPLICANT: PIENENBROCK, Christian  
; APPLICANT: BERLIN, Kurt  
; TITLE OF INVENTION: Diagnosis of Diseases Associated with Gene Regulation

```

: FILE REFERENCE: 5013.1003
: CURRENT APPLICATION NUMBER: US/10/239, 676
: CURRENT FILING DATE: 2002-09-24
: PRIOR APPLICATION NUMBER: PCT/EP01/03968
  DE 10019058.8
  DE 10019173.8
  DE 10032529.7
  DE 10043826.1
: PRIOR FILING DATE: 2001-04-06
  2000-04-06
  2000-04-07
  2000-05-30
  2000-09-01
: NUMBER OF SEQ ID NOS: 228
  SEQ ID NO 209
  LENGTH: 11812
  TYPE: DNA
: ORGANISM: Artificial Sequence
: FEATURE:
: OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens
US-10-239-676-209

```

Query Match	3.7%	Score 42.8	DB 9	Length 1812
Best Local Similarity	45.6%	Pred. No. 3.1		
Matches .187	Conservative	0	Mismatches 222	Indels 1
			Gaps	1

Oy	633	TTTATGCAATTTATTTGATTTGTAAGACCAATAAAATTTGCGAAACATAAATCTTT	692
Dd	4307	TTTATGCTATGTTTTTTGTTGTAAGCTTTAAATATATGTTTTATTTTGCGTTTACT	4366
Oy	693	AAATATTTCTATTCCTGTACACACCTTTTTCAGACGCTCCAGTTGGAGCATTTGCCCTC	752
Dd	4367	TTAATATATTTTTTTTTTATATATAGTTTTTATGTAATTTGTAATTTATATAGATTTTT	4426
Oy	753	TAAAGCTTTTAAACATATTTTTTACATGCTTATATACATTTCCACTAAATGACATTTCT	811
Dd	4427	TTTATATTTTTTAAATTTTATTTGAGTTTTTTTAAATATTTTAAATTAATAGGCTTAAT	4486
Oy	812	ATAATAAATTTAAACACATGATCTTGGTACTACATACACTCAGTGAACCCAGCCTATG	871
Dd	4487	TTTATATTTTTTAAATTTATAGTTTTTATATATAATTTGAAATTTGTTTTATTTATTTTT	4546
Oy	872	CAAAATATAAATCTTTTATATATATTATCTATGAGATCGACACAAATATAACACTG	931
Dd	4547	TGTTTAAATAATTTATTTTATTTTATATATGCTTTATAAAATGATTAACGCTATGTAAT	4606
Oy	932	GAGAAGTGAAGCTTTTGGTTATTTAGGTTAATTTCTAGTAAACACAGTTGCCTGTTTT	991
Dd	4607	GAAAAGCTTTAGATATTTTTTTTTTGTATGATTAATTTTTTGGAAAGTAGGCGTTTTTTATG	4666
Oy	992	CAGTTAAACAGGTATAGCCATTTTAAATATATGCTTTTCAAAACACTT	1041
Dd	4667	TTGTTTTTGCTGTAATTTAANGTAGTATATGTTTAGATTAATTAATTT	4716

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

US-10-172-086-18

Query Match	3.7%	Score 42.8;	DB 9;	Length 18997;
Best Local Similarity	47.7%	Pred. No. 4;		
Matches 125; Conservative	0;	Mismatches 137;	Indels 0;	Gaps 0;

[illegible]

Search completed: June 28, 2003, 03:41:55  
Job time : 191.807 secs

GENOME  
RESEARCH



Herculase™

HOME HELP FEEDBACK SUBSCRIPTIONS ARCHIVE SEARCH TABLE OF CONTENTS

Institution: US PATENT & TRADEMARK OFFICE Sign In as Individual

Genome Research, Vol 6, 791-806, Copyright © 1996 by Cold Spring Harbor Laboratory Press

## ARTICLES


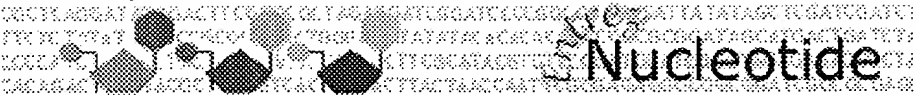
# Normalization and subtraction: two approaches to facilitate gene discovery

MF Bonaldo, G Lennon and MB Soares

Department of Psychiatry, College of Physicians and Surgeons of Columbia University, New York, New York, USA.

- ▶ [Email this article to a friend](#)
- ▶ Similar articles found in:  
     [Genome Online](#)  
     [PubMed](#)
- ▶ [PubMed Citation](#)
- ▶ This Article has been cited by:  
     [other online articles](#)
- ▶ Search PubMed for articles by:  
     [Bonaldo, M. F.](#) || [Soares, M. B.](#)
- ▶ Alert me when:  
     [new articles cite this article](#)
- ▶ [Download to Citation Manager](#)

Large-scale sequencing of cDNAs randomly picked from libraries has proven to be a very powerful approach to discover (putatively) expressed sequences that, in turn, once mapped, may greatly expedite the process involved in the identification and cloning of human disease genes. However, the integrity of the data and the pace at which novel sequences can be identified depends to a great extent on the cDNA libraries that are used. Because altogether, in a typical cell, the mRNAs of the prevalent and intermediate frequency classes comprise as much as 50-65% of the total mRNA mass, but represent no more than 1000- 2000 different mRNAs, redundant identification of mRNAs of these two frequency classes is destined to become overwhelming relatively early in any such random gene discovery programs, thus seriously compromising their cost-effectiveness. With the goal of facilitating such efforts, previously we developed a method to construct directionally cloned normalized cDNA libraries and applied it to generate infant brain (INIB) and fetal liver/spleen (INFLS) libraries, from which a total of 45,192 and 86,088 expressed sequence tags, respectively, have been derived. While improving the representation of the longest cDNAs in our libraries, we developed three additional methods to normalize cDNA libraries and generated over 35 libraries, most of which have been contributed to our integrated Molecular Analysis of Genomes and Their Expression (IMAGE) Consortium and thus distributed widely and used for sequencing and mapping. In an attempt to facilitate the process of gene discovery further, we have also developed a subtractive hybridization approach designed specifically to eliminate (or reduce significantly the representation of) large pools of arrayed and (mostly) sequenced clones from normalized libraries yet to be (or just partly) surveyed. Here we present a detailed description and a comparative analysis of four methods that we developed and used to generate normalize cDNA libraries from human (15), mouse (3), rat (2), as well as the parasite *Schistosoma mansoni* (1). In addition, we describe the construction and preliminary characterization of a subtracted liver/spleen library (INFLS-SI) that resulted from the elimination (or reduction of representation) of -5000 INFLS-IMAGE clones from the INFLS library.

  **Nucleotide**

PubMed Nucleotide Protein Genome Structure PMC Taxonomy OMIM Bio

Search  for

Limits Preview/Index History Clipboard Details

Show:

☐ 1: AW274482.xv30c09.x1 Soares...[gi:6661512]

[Links](#)

#### IDENTIFIERS

dbEST Id: 3657583  
EST name: xv30c09.x1  
GenBank Acc: AW274482  
GenBank gi: 6661512

#### CLONE INFO

Clone Id: IMAGE:2814640 (3')  
Source: NCI  
DNA type: cDNA

#### PRIMERS

Sequencing: -40UP from Gibco  
PolyA Tail: Unknown

#### SEQUENCE

```
AGAAAAATAGAAAAGTTTACTACTTTGAAAAGGAACTATGACAAACAAGTATATATTCA
GGAAAGGGACTCCTAGAACTTGAGCAACAAAAAGAGTTCAACCTCGGCACAGTGGAGTC
TGCAGTGTCTCCGCAACAACAGAAATGAACCTGTGTTCTCACATCCTTGAGAGTCGAGG
TTTGCATTTCTCTACAGAGCTAGGGCTGAATGAAGCAAAGAGACCTCTGCCACCCCACT
AGCTGGCCCAGTCCTGGAAGCGGAAGCAGTCACTTGGCATCTTCCACACTGTGTTGCTGG
GTGAGGCCTGGGCGGT CAGGATGAAGTTCTGGTTGAAGTCCCGTTGTTTGTTCCTCAA
ACTTCACTGATCCACAGATGACAACAAGGACCGTGGTCTGGCTTGGTGTGGCTTCATCAT
GAACAGGCTGGCAGTCTACCACGCTGATTTGGAACGCTGGAAGGCAACATTTCAAAA
ACTCACTCAAGGATTCTTGTCTGAAACAGCATTGCCATTCCAGACCAGGTGGCTGTGC
CCATGTACAGGCGGGACAGCAAACGCCGCGCTTATCCATGGTGGTGTAGTAGACATTGA
CAGACTCCTCAGCAGCTCTGCAGGCCTGATC
```

Quality: High quality sequence stops at base: 466

Entry Created: Jan 3 2000  
Last Updated: Jan 3 2000

#### COMMENTS

This clone is available royalty-free through LLNL ; contact the IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.

PUTATIVE ID Assigned by submitter  
WP:Y71F9A\_279.B CE22843 ;

#### LIBRARY

Lib Name: Soares\_NFL\_T\_GBC\_S1  
Organism: Homo sapiens  
Organ: pooled  
Lab host: DH10B  
Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker

h      cb      hg   e   e   e   fcg   c      e   e   e   b   c   e   e

R. Site 1: Not I  
R. Site 2: Eco RI  
Description: Equal amounts of plasmid DNA from three normalized libraries (fetal lung NbHL19W, testis NHT, and B-cell NCI\_CGAP\_GCB1) were mixed, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from pools of 5,000 clones made from the same 3 libraries. The pools consisted of I.M.A.G.E. clones 297480-302087, 682632-687239, 726408-728711, and 729096-731399. Subtraction by Bento Soares and M. Fatima Bonaldo.

**SUBMITTER**

Name: Robert Strausberg, Ph.D.  
E-mail: [cgapbs-r@mail.nih.gov](mailto:cgapbs-r@mail.nih.gov)

**CITATIONS**

Title: National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index  
Authors: NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>  
Year: 1997  
Status: Unpublished

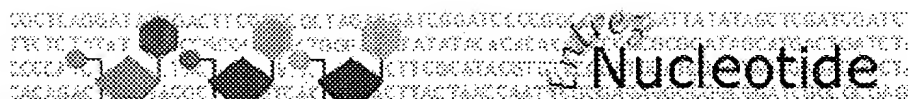
**MAP DATA**

---

[Disclaimer](#) | [Write to the Help Desk](#)  
[NCBI](#) | [NLM](#) | [NIH](#)

10/30/2003 12:44:50





PubMed

Nucleotide

Protein

Genome

Structure

PMC

Taxonomy

OMIM

Bio

Search  for 

Go

Clear

Limits

Preview/Index

History

Clipboard

Details

Display

Show:

Send to

☐ 1: BF195575. 7n85b11.x1 NCI\_CG...[gi:11082605]

Links

## IDENTIFIERS

dbEST Id: 6660525  
EST name: 7n85b11.x1  
GenBank Acc: BF195575  
GenBank gi: 11082605

## CLONE INFO

Clone Id: IMAGE:3571173 (3')  
Source: NCI  
DNA type: cDNA

## PRIMERS

PolyA Tail: Unknown

## SEQUENCE

AAGAAGCAACATGTGTATTATTAAATTATTTGTCAGAATTTCCAGAATCAGAGTCTCTAC  
TGGGCAAGTAGAAAAATAGAAAAGTTTACTACTTTGAAAAGGAACTATGACAAACAAGT  
ATATATTAGGAAAGGGACTCCTAGAACTTGAGCAACAAAAAGAGTTCAACCTCGGCAC  
AGTGGAGTCTGCAGTGTCTCCGCAACAACAGAAATGAAGTGTGTTCCCTCACATCCTGA  
GAGTCGAGGTTTGCATTTCTCTACAGAGCTAGGGCTGAATGAAGCAAAGAGACCTCTGCC  
ACCCCCACTAGCTGGCCAGTCCTGGAAGCGGAAGCAGTCACTTGCATCTTCCCACTG  
TGTTGCTGGGTGAGGCCTGNGCGGTGAGGATGAAGTTCTGGTTGAAGTCCCCTGTTTGT  
TCCCCTCAAACCTTCACTGATCCACAGATGACAACAAGGACCGTGGTCTGGCTTGGTGTGG  
CTTCATCATGAACAGGCCTGCAGTCTACCACGCTGATTTGGAAGTCCGCTGGAAGGCAACA  
TTTCAAAAACTCACTCAAGGATTCTTGTGCTGAAACAGCATTGCCATTCAGAACCAAGG  
GGTGCTGTGCCCATGTACAGGCCGGACAGCAAACGCCCGCGCTTATCCCTGGTGGTGTG  
TGACATTGACAACTTCTCACAGCTCTGCAAGGCTGATCCACATAGTTGTGAAATCCCAA

Quality: High quality sequence stops at base: 466

Entry Created: Nov 3 2000  
Last Updated: Nov 3 2000

## COMMENTS

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D.,  
Michael R. Emmert-Buck, M.D., Ph.D. cDNA Library  
Preparation: M. Bento Soares, Ph.D. cDNA Library Arrayed by:  
Christa Prange, The I.M.A.G.E. Consortium DNA Sequencing by:  
Washington University Genome Sequencing Center  
Clone distribution: NCI-CGAP clone distribution information  
can be found through the I.M.A.G.E. Consortium/LLNL, send  
email to: [info@image.llnl.gov](mailto:info@image.llnl.gov)

## PUTATIVE ID

Assigned by submitter  
TR:Q9UUK6 Q9UUK6 NTF2-RELATED EXPORT PROTEIN NXT1..;

## LIBRARY

h cb hg e e e fcg c e e e b c e e

Lib Name: NCI\_CGAP\_Ov18  
Organism: Homo sapiens  
Organ: ovary  
Tissue type: fibrotheoma  
Lab host: DH10B (phage-resistant)  
Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker  
R. Site 1: Not I  
R. Site 2: Eco RI  
Description: 1st strand cDNA was primed with a Not I - oligo(dT) primer  
[5' TGTTACCAATCTGAAGTGGGAGCGGCCGCGACATTTTTTTTTTTTTTTTTT  
3']; double-stranded cDNA was ligated to Eco RI adaptors  
(Pharmacia), digested with Not I and cloned into the Not I  
and Eco RI sites of the modified pT7T3 vector. Library went  
through one round of normalization, and was constructed by  
Bento Soares and M. Fatima Bonaldo.

**SUBMITTER**

Name: Robert Strausberg, Ph.D.  
E-mail: cgapbs-r@mail.nih.gov

**CITATIONS**

Title: National Cancer Institute, Cancer Genome Anatomy Project  
(CGAP), Tumor Gene Index  
Authors: NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>  
Year: 1997  
Status: Unpublished

**MAP DATA**

---

[Disclaimer](#) | [Write to the Help Desk](#)  
[NCBI](#) | [NLM](#) | [NIH](#)

Jul 30 2003 12:44:50